[redacted] Search results for 08/494,282

```
show files
File 155:MEDLINE(R) 1966-2000/Dec W4
         (c) format only 2000 Dialog Corporation
File   2:INSPEC 1969-2001/Mar W1
         (c) 2001 Institution of Electrical Engineers
File   8:Ei Compendex(R) 1970-2001/Feb W2
         (c) 2001 Engineering Info. Inc.
File  65:Inside Conferences 1993-2001/Feb W4
         (c) 2001 BLDSC all rts. reserv.
File 144:Pascal 1973-2001/Mar W1
         (c) 2001 INIST/CNRS
File  35:Dissertation Abstracts Online 1861-2000/Dec
         (c) 2000 UMI
File 239:Mathsci 1940-2001/Mar
         (c) 2001 American Mathematical Society
File   5:Biosis Previews(R) 1969-2001/Feb W4
         (c) 2001 BIOSIS
File  73:EMBASE 1974-2001/Feb W4
         (c) 2001 Elsevier Science B.V.
File 151:HealthSTAR 1975-2000/Dec
         (c) format only 2000 The Dialog Corporation
File  77:Conference Papers Index 1973-2001/Mar
         (c) 2001 Cambridge Sci Abs
File  76:Life Sciences Collection 1982-2001/Dec
         (c) 2001 Cambridge Sci Abs
File 351:Derwent WPI 1963-2000/UD,UM &UP=200111
         (c) 2001 Derwent Info Ltd
File 202:Information Science Abs. 1966-2000/ISSUE 09
         (c) Information Today, Inc
File 238:Abs. in New Tech & Eng. 1981-2001/Feb
         (c) 2001 Reed-Elsevier (UK) Ltd.
?ds

Set     Items   Description
S1      46321   ((GENETIC? ? OR EVOLUTION?) (3N) ALGORITHM? ?)
S2       1933   S1 AND (NUCLEIC OR DNA OR RIBONUCLEIC OR RNA OR DEOXYRIBON-
            UCLEIC OR PROTEIN? ? OR PEPTIDE? ?)
S3         69   S1 AND ((MODIF? OR GENERAT?)(5N) (NUCLEIC OR DNA OR RIBONU-
           CLEIC OR RNA OR DEOXYRIBONUCLEIC OR PROTEIN? ? OR PEPTIDE? ?))
S4         38   RD S3 (unique items)
S5        107   S1 AND (DESIGN? (5N) (NUCLEIC OR DNA OR RIBONUCLEIC OR RNA
           OR DEOXYRIBONUCLEIC OR PROTEIN? ? OR PEPTIDE? ?))
S6        137   S4 OR S5
S7         87   RD S6 (unique items)
S8        103   AU=SELIFONOV S? OR AU=SELIFONOV, S?
S9        212   AU=STEMMER W? OR AU=STEMMER, W?
S10       262   AU=GUSTAFSSON C? OR AU=GUSTAFSSON, C?
S11      1711   AU= TOBIN M? OR AU=TOBIN, M?
S12      2275   S8-S11
S13         5   S12 AND S2
S14         2   RD S13 (unique items)
S15        88   S7 OR S14
?t 15/7/all

 15/7/1     (Item 1 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.
```

10464345 20336890
Rational design of faster associating and tighter binding protein complexes.
Selzer T; Albeck S; Schreiber G
Weizmann Institute of Science, Department of Biological Chemistry, Rehovot, 76100 Israel.
Nature structural biology (UNITED STATES) Jul 2000, 7 (7) p537-41, ISSN 1072-8368 Journal Code: B98
Languages: ENGLISH
Document type: JOURNAL ARTICLE
A protein design strategy was developed to specifically enhance the rate of association (k(on)) between a pair of proteins without affecting the rate of dissociation (k(off)). The method is based on increasing the electrostatic attraction between the proteins by incorporating charged residues in the vicinity of the binding interface. The contribution of mutations towards the rate of association was calculated using a newly developed computer algorithm, which predicted accurately the rate of association of mutant protein complexes relative to the wild type. Using this design strategy, the rate of association and the affinity between TEM1 beta-lactamase and its protein inhibitor BLIP was enhanced 250-fold, while the dissociation rate constant was unchanged. The results emphasize that long range electrostatic forces specifically alter k(on), but do not effect k(off). The design strategy presented here is applicable for increasing rates of association and affinities of protein complexes in general.

15/7/2 (Item 2 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10449306 20296807
Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design.
Voigt CA; Gordon DB; Mayo SL
Biochemistry Option Divisions of Biology and Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125, USA.
Journal of molecular biology (ENGLAND) Jun 9 2000, 299 (3) p789-803, ISSN 0022-2836 Journal Code: J6V
Contract/Grant No.: GM 08346, GM, NIGMS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Finding the minimum energy amino acid side-chain conformation is a fundamental problem in both homology modeling and protein design . To address this issue, numerous computational algorithms have been proposed. However, there have been few quantitative comparisons between methods and there is very little general understanding of the types of problems that are appropriate for each algorithm. Here, we study four common search techniques: Monte Carlo (MC) and Monte Carlo plus quench (MCQ); genetic algorithms (GA); self-consistent mean field (SCMF); and dead-end elimination (DEE). Both SCMF and DEE are deterministic, and if DEE converges, it is guaranteed that its solution is the global minimum energy conformation (GMEC). This provides a means to compare the accuracy of SCMF and the stochastic methods. For the side-chain placement calculations, we find that DEE rapidly converges to the GMEC in all the test cases. The other algorithms converge on significantly incorrect solutions; the average fraction of incorrect rotamers for SCMF is 0.12, GA 0.09, and MCQ 0.05. For

the protein design calculations, design positions are progressively added to the side-chain placement calculation until the time required for DEE diverges sharply. As the complexity of the problem increases, the accuracy of each method is determined so that the results can be extrapolated into the region where DEE is no longer tractable. We find that both SCMF and MCQ perform reasonably well on core calculations (fraction amino acids incorrect is SCMF 0.07, MCQ 0.04), but fail considerably on the boundary (SCMF 0.28, MCQ 0.32) and surface calculations (SCMF 0.37, MCQ 0.44). Copyright 2000 Academic Press.

15/7/3 (Item 3 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10403295 20237724
An algorithm for the prediction of proteasomal cleavages.
Kuttler C; Nussbaum AK; Dick TP; Rammensee HG; Schild H; Hadeler KP
Biomathematik, University of Tubingen, Auf der Morgenstelle 10, Tubingen, D-72076, Germany.
Journal of molecular biology (ENGLAND) May 5 2000, 298 (3) p417-29, ISSN 0022-2836 Journal Code: J6V
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Proteasomes, major proteolytic sites in eukaryotic cells, play an important part in major histocompatibility class I (MHC I) ligand generation and thus in the regulation of specific immune responses. Their cleavage specificity is of outstanding interest for this process.In order to generalize previously determined cleavage motifs of 20 S proteasomes, we developed network-based model proteasomes trained by an evolutionary algorithm with experimental cleavage data of yeast and human 20 S proteasomes. A window of ten flanking amino acid residues proved sufficient for the model proteasomes to reproduce the experimental results with 98-100 % accuracy. Actual experimental data were reproduced significantly better than randomly selected cleavage sites, suggesting that our model proteasomes were able to extract rules inherent to proteasomal cleavage data. The affinity parameters of the model, which decide for or against cleavage, correspond with the cleavage motifs determined experimentally. The predictive power of the model was verified for unknown (to the program) test conditions: the prediction of cleavage numbers in proteins and the generation of MHC I ligands from short peptides.In summary, our model proteasomes reproduce and predict proteasomal cleavages with high degree of accuracy. They present a promising approach for predicting proteasomal cleavage products in future attempts and, in combination with existing algorithms for MHC I ligand prediction, will be tested to improve cytotoxic T lymphocyte epitope prediction. Copyright 2000 Academic Press.

15/7/4 (Item 4 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10242853 20054763
Use of a quantitative structure-property relationship to design larger model proteins that fold rapidly.
Dinner AR; Verosub E; Karplus M
Department of Chemistry and Chemical Biology and Committee on Higher

Degrees in Biophysics, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA and Laboratoire de Chimie Biophysique, Institut le Bel, Universite Louis Pasteur, 4.

Protein engineering (ENGLAND) Nov 1999, 12 (11) p909-17, ISSN 0269-2139 Journal Code: PR1

Languages: ENGLISH

Document type: JOURNAL ARTICLE

A quantitative structure-property relationship (QSPR) was used to design model protein sequences that fold repeatedly and relatively rapidly to stable target structures. The specific model was a 125-residue heteropolymer chain subject to Monte Carlo dynamics on a simple cubic lattice. The QSPR was derived from an analysis of a database of 200 sequences by a statistical method that uses a genetic algorithm to select the sequence attributes that are most important for folding and a neural network to determine the corresponding functional dependence of folding ability on the chosen attributes. The QSPR depends on the number of anti-parallel sheet contacts, the energy gap between the native state and quasi-continuous part of the spectrum and the total energy of the contacts between surface residues. Two Monte Carlo procedures were used in series to optimize both the target structures and the sequences. We generated 20 fully optimized sequences and 60 partially optimized control sequences and tested each for its ability to fold in dynamic MC simulations. Although sequences in which either the number of anti-parallel sheet contacts or the energy of the surface residues is non-optimal are capable of folding almost as well as fully optimized ones, sequences in which only the energy gap is optimized fold markedly more slowly. Implications of the results for the design of proteins are discussed.

15/7/5 (Item 5 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10235389 20057857

Tendency for local repetitiveness in amino acid usages in modern proteins.

Nishizawa K; Nishizawa M; Kim KS

Department of Biochemistry, Teikyo University School of Medicine, Kaga, Itabashi, Tokyo, 173, Japan. kazunet@med.teikyo-u.ac.jp

Journal of molecular biology (ENGLAND) Dec 10 1999, 294 (4) p937-53, ISSN 0022-2836 Journal Code: J6V

Languages: ENGLISH

Document type: JOURNAL ARTICLE

Systematic analyses of human proteins show that neural and immune system-specific, and therefore, relatively "modern" proteins have a tendency for repetitive use of amino acids at a local scale (approximately 1-20 residues), while ancient proteins (human homologues of Escherichia coli proteins) do not. Those protein subsegments which are unique based on homology search account for the repetitiveness. Simulation shows that such repetitiveness can be maintained by frequent duplication on a very short scale (one to two codons) in the presence of substitutive point mutation, while the latter tends to mitigate the repetitiveness. DNA analyses also show the presence of cryptic (i.e. "out of the codon frame") repetitiveness, which cannot fully be explained by features in protein sequences. Simulative modification of the amino acid sequences of immune system-specific proteins estimate that 2.4 duplication events occur during the period equivalent to ten events of substitution mutation. It is also

suggested that the repetitiveness leads to longitudinal unevenness within a given peptide domain. Those peptide motifs which contain similarly charged residues are likely to be generated more frequently in the presence of the tendency for repetitiveness than in its absence. Therefore, the neutral propensity of DNA for duplication, which can also tend to generate repetitiveness in amino acid sequences, seems to be manifested primarily when the constraints on amino acid sequences are relatively weak, and yet may be positively contributing to generation of unevenness in modern proteins . Copyright 1999 Academic Press.

15/7/6 (Item 6 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10223253 20047399
Efficient algorithms for protein sequence design and the analysis of certain evolutionary fitness landscapes.
Kleinberg JM
Department of Computer Science, Cornell University, Ithaca, New York 14853, USA. kleinber@cs.cornell.edu
Journal of computational biology (UNITED STATES) Fall-Winter 1999, 6 (3-4) p387-404, ISSN 1066-5277 Journal Code: CGW
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Protein sequence design is a natural inverse problem to protein structure prediction: given a target structure in three dimensions, we wish to design an amino acid sequence that is likely fold to it. A model of Sun, Brem, Chan, and Dill casts this problem as an optimization on a space of sequences of hydrophobic (H) and polar (P) monomers; the goal is to find a sequence that achieves a dense hydrophobic core with few solvent-exposed hydrophobic residues. Sun et al. developed a heuristic method to search the space of sequences, without a guarantee of optimality or near-optimality; Hart subsequently raised the computational tractability of constructing an optimal sequence in this model as an open question. Here we resolve this question by providing an efficient algorithm to construct optimal sequences; our algorithm has a polynomial running time, and performs very efficiently in practice. We illustrate the implementation of our method on structures drawn from the Protein Data Bank. We also consider extensions of the model to larger amino acid alphabets, as a way to overcome the limitations of the binary H/P alphabet. We show that for a natural class of arbitrarily large alphabets, it remains possible to design optimal sequences efficiently. Finally, we analyze some of the consequences of this sequence design model for the study of evolutionary fitness landscapes. A given target structure may have many sequences that are optimal in the model of Sun et al.; following a notion raised by the work of J. Maynard Smith, we can ask whether these optimal sequences are "connected" by successive point mutations. We provide a polynomial-time algorithm to decide this connectedness property, relative to a given target structure. We develop the algorithm by first solving an analogous problem expressed in terms of submodular functions, a fundamental object of study in combinatorial optimization.

15/7/7 (Item 7 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10172868 99400913
Automated docking of peptides and proteins by using a genetic algorithm combined with a tabu search.
Hou T; Wang J; Chen L; Xu X
Department of Chemistry, Peking University Jiuyuan Molecular Design Laboratory and Department of Technical Physics, Peking University, Beijing 100871, China.
Protein engineering (ENGLAND) Aug 1999, 12 (8) p639-48, ISSN 0269-2139 Journal Code: PR1
Languages: ENGLISH
Document type: JOURNAL ARTICLE
A genetic algorithm (GA) combined with a tabu search (TA) has been applied as a minimization method to rake the appropriate associated sites for some biomolecular systems. In our docking procedure, surface complementarity and energetic complementarity of a ligand with its receptor have been considered separately in a two-stage docking method. The first stage was to find a set of potential associated sites mainly based on surface complementarity using a genetic algorithm combined with a tabu search. This step corresponds with the process of finding the potential binding sites where pharmacophores will bind. In the second stage, several hundreds of GA minimization steps were performed for each associated site derived from the first stage mainly based on the energetic complementarity. After calculations for both of the two stages, we can offer several solutions of associated sites for every complex. In this paper, seven biomolecular systems, including five bound complexes and two unbound complexes, were chosen from the Protein Data Bank (PDB) to test our method. The calculated results were very encouraging-the hybrid minimization algorithm successfully reaches the correct solutions near the best binded modes for these protein complexes. The docking results not only predict the bound complexes very well, but also get a relatively accurate complexed conformation for unbound systems. For the five bound complexes, the results show that surface complementarity is enough to find the precise binding modes, the top solution from the tabu list generally corresponds to the correct binding mode. For the two unbound complexes, due to the conformational changes upon binding, it seems more difficult to get their correct binding conformations. The predicted results show that the correct binding mode also corresponds to a relatively large surface complementarity score. In these two test cases, the correct solution can be found in the top several solutions from the tabu list. For unbound complexes, the interaction energy from energetic complementarity is very important, it can be used to filter these solutions from the surface complementarity. After the evaluation of the energetic complementarity, the conformations and orientations close to the crystallographically determined structures are resolved. In most cases, the smallest root mean square distance (r.m.s.d.) from the GA combined with TA solutions is in a relatively small region. Our program of automatic docking is really a universal one among the procedures used for the theoretical study of molecular recognition.

15/7/8 (Item 8 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10134665 99318885
Side-chain and backbone flexibility in protein core design.
Desjarlais JR; Handel TM

Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720, USA.
Journal of molecular biology (ENGLAND) Jul 2 1999, 290 (1) p305-18, ISSN 0022-2836 Journal Code: J6V
Languages: ENGLISH
Document type: JOURNAL ARTICLE
We have developed a computational approach for the design and prediction of hydrophobic cores that includes explicit backbone flexibility. The program consists of a two-stage combination of a genetic algorithm and monte carlo sampling using a torsional model of the protein. Backbone structures are evaluated either by a canonical force-field or a constraining potential that emphasizes the preservation of local geometry. The utility of the method for protein design and engineering is explored by designing three novel hydrophobic core variants of the protein 434 cro. We use the new method to evaluate these and previously designed 434 cro variants, as well as a series of phage T4 lysozyme variants. In order to properly evaluate the influence of backbone flexibility, we have also analyzed the effects of varying amounts of side-chain flexibility on the performance of fixed backbone methods. Comparison of results using a fixed versus flexible backbone reveals that, surprisingly, the two methods are almost equivalent in their abilities to predict relative experimental stabilities, but only when full side-chain flexibility is allowed. The prediction of core side-chain structure can vary dramatically between methods. In some, but not all, cases the flexible backbone method is a better predictor of structure. The development of a flexible backbone approach to core design is particularly important for attempts at de novo protein design , where there is no prior knowledge of a precise backbone structure. Copyright 1999 Academic Press.

15/7/9 (Item 9 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10119206 99063705
SCOP: a Structural Classification of Proteins database.
Hubbard TJ; Ailey B; Brenner SE; Murzin AG; Chothia C
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.
Nucleic acids research (ENGLAND) Jan 1 1999, 27 (1) p254-6, ISSN 0305-1048 Journal Code: O8L
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The Structural Classification of Proteins (SCOP) database provides a detailed and comprehensive description of the relationships of all known proteins structures. The classification is on hierarchical levels: the first two levels, family and superfamily, describe near and far evolutionary relationships; the third, fold, describes geometrical relationships. The distinction between evolutionary relationships and those that arise from the physics and chemistry of proteins is a feature that is unique to this database, so far. The database can be used as a source of data to calibrate sequence search algorithms and for the generation of population statistics on protein structures. The database and its associated files are freely accessible from a number of WWW sites mirrored from URL http://scop. mrc-lmb.cam.ac.uk/scop/

15/7/10 (Item 10 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

10069219 99377180
Origin and properties of non-coding ORFs in the yeast genome.
Mackiewicz P; Kowalczuk M; Gierlik A; Dudek MR; Cebrat S
Institute of Microbiology, Wroclaw University, ul. Przybyszewskiego 63/77, 54-148 Wroclaw, Poland.
Nucleic acids research (ENGLAND) Sep 1 1999, 27 (17) p3503-9, ISSN 0305-1048 Journal Code: O8L
Languages: ENGLISH
Document type: JOURNAL ARTICLE
In a recent paper we have estimated the total number of protein coding open reading frames (ORFs) in the Saccharomyces cerevisiae genome, based on their properties, at about 4800. This number is much smaller than the 5800-6000 which is widely accepted. In this paper we analyse differences between the set of ORFs with known phenotypes annotated in the Munich Information Centre for Protein Sequences (MIPS) database and ORFs for which the probability of coding, counted by us, is very low. We have found that many of the latter ORFs have properties of antisense sequences of coding ORFs, which suggests that they could have been generated by duplication of coding sequences. Since coding sequences generate ORFs inside themselves, with especially high frequency in the antisense sequences, we have looked for homology between known proteins and hypothetical polypeptides generated by ORFs under consideration in all the six phases. For many ORFs we have found paralogues and orthologues in phases different than the phase which had been assumed in the MIPS database as coding.

15/7/11 (Item 11 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09810038 99095575
Patenting computer- designed peptides.
Patel S; Stott IP; Bhakoo M; Elliott P
Unilever Research, Port Sunlight Laboratory, Wirral, U.K.
Journal of computer-aided molecular design (NETHERLANDS) Nov 1998, 12 (6) p543-56, ISSN 0920-654X Journal Code: JCB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The problem of designing new peptides that possess specific properties, such as bactericidal activity, is of wide interest. Recently, attention has focused on the use of Computer-Aided Molecular Design techniques in parallel with more traditional 'synthesise and test' methods. These techniques may typically use Genetic Algorithms to optimise' molecules based on Neural Network models that predict activity. In this paper we describe a successful application of this Molecular Design methodology that has resulted in novel bactericidal peptides of real value. A key issue for commercial utilisation of such results is the ability to protect the intellectual property rights associated with the discovery of new molecules. Typically peptide patents use structural templates of amino acid hydrophobicity-hydrophilicity that define highly regular peptide patent spaces. In an extension of established patenting practice we describe a patent application that uses a Neural Net predictive model to define the regions of peptide space that we claim within the patent. This

formalism makes no a priori assumptions about the regularity of the patent space. A preliminary comparative investigation of the shape and size of this and other bactericidal peptide patent spaces is conducted.

15/7/12 (Item 12 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09673442 98390143
Applications of genetic algorithms in molecular diversity.
Weber L
Hoffmann-La Roche AG, Basel, Switzerland. lutz.weber@roche.com
Current opinion in chemical biology (ENGLAND) Jun 1998, 2 (3) p381-5, ISSN 1367-5931 Journal Code: C4U
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, TUTORIAL
The definition of molecular diversity and the development of measures for assessing the similarity or dissimilarity of molecules are central tasks for the design of novel biologically active compounds. Combinatorial chemistry allows the coupling of mathematical optimisation methods that do not require the a priori knowledge of structure-activity relationships with the synthesis of biologically active compounds. Genetic algorithms that computationally mimic Darwinian evolution have proven to be useful in solving multidimensional problems and are now being used successfully in various areas of combinatorial chemistry. Applications have been developed that help in the selection of diverse compound libraries and in the synthesis of biologically active molecules. (28 Refs.)

15/7/13 (Item 13 from file: 155)
DIALOG(R)File 155:MEDLINE(R)
(c) format only 2000 Dialog Corporation. All rts. reserv.

09672412 98362517
Focus-2D: a new approach to the design of targeted combinatorial chemical libraries.
Cho SJ; Zheng W; Tropsha A
Laboratory for Molecular Modeling, School of Pharmacy, University of North Carolina, Chapel Hill 27599-7360, USA.
Pacific Symposium on Biocomputing (SINGAPORE) 1998, p305-16, Journal Code: CWQ
Contract/Grant No.: MH 40537, MH, NIMH; HD03310, HD, NICHD; MH33127, MH, NIMH
Languages: ENGLISH
Document type: JOURNAL ARTICLE
A strategy for rational design of targeted combinatorial libraries is described. The aim of this approach is to select a subset of available building blocks for the library synthesis that are most likely to be present in the active compounds. Building blocks that are used in the underlying combinatorial chemical reaction are randomly assembled to produce virtual combinatorial library compounds, which are represented by various chemical descriptors. Stochastic algorithms (simulated annealing, genetic algorithms , neural net methods) are used to search the potentially large structural space of virtual chemical libraries in order to identify compounds similar to lead compound(-s). The selection of a virtual molecule as a candidate for the targeted library is based either on

its chemical similarity to a biologically active probe or on its biological activity predicted from a pre-constructed QSAR equation. Frequency analysis of building block composition of the selected virtual compounds identifies building blocks that can be used in combinatorial synthesis of chemical libraries with high similarity to the lead compound(-s). This method is applied to rational design of the library with bradykinin potentiating activity. Twenty eight bradykinin potentiating pentapeptides were used as a training set for the development of a QSAR equation, and, alternatively, two active pentapeptides, VEWAK and VKWAP, were used as probe molecules. In each case, the frequency distribution of amino acids in the top 100 peptides suggested by the method resembles the frequency distribution of amino acids found in the active peptides. The results obtained after GA optimization also compared favorably with those obtained by the exhaustive analysis of all possible 3.2 millions pentapeptides.

15/7/14 (Item 14 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09669596 98399438
Computer search algorithms in protein modification and design.
Desjarlais JR; Clarke ND
Department of Chemistry, Pennsylvania State University, University Park 16802, USA. jrd@chem.psu.edu
Current opinion in structural biology (ENGLAND) Aug 1998, 8 (4) p471-5 , ISSN 0959-440X Journal Code: B9V
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, TUTORIAL
The computer-aided design of protein sequences requires efficient search algorithms to handle the enormous combinatorial complexity involved. A variety of different algorithms have now been applied with some success. The choice of algorithm can influence the representation of the problem in several important ways--the discreteness of the configuration, the types of energy terms that can be used and the ability to find the global minimum energy configuration. The use of dead end elimination to design the complete sequence for a small protein motif and the use of genetic and mean-field algorithms to design hydrophobic cores for proteins represent the major themes of the past year. (40 Refs.)

15/7/15 (Item 15 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09590675 98365504
From synthetic coiled coils to functional proteins: automated design of a receptor for the calmodulin-binding domain of calcineurin.
Ghirlanda G; Lear JD; Lombardi A; DeGrado WF
Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, 19104-6059, USA.
Journal of molecular biology (ENGLAND) Aug 14 1998, 281 (2) p379-91, ISSN 0022-2836 Journal Code: J6V
Contract/Grant No.: GM54616, GM, NIGMS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
A series of synthetic receptors capable of binding to the

calmodulin-binding domain of calcineurin (CN393-414) was designed, synthesized and characterized. The design was accomplished by docking CN393-414 against a two-helix receptor, using an idealized three-stranded coiled coil as a starting geometry. The sequence of the receptor was chosen using a side-chain re-packing program, which employed a genetic algorithm to select potential binders from a total of 7.5x10(6) possible sequences. A total of 25 receptors were prepared, representing 13 sequences predicted by the algorithm as well as 12 related sequences that were not predicted. The receptors were characterized by CD spectroscopy, analytical ultracentrifugation, and binding assays. The receptors predicted by the algorithm bound CN393-414 with apparent dissociation constants ranging from 0.2 microM to >50 microM. Many of the receptors that were not predicted by the algorithm also bound to CN393-414. Methods to circumvent this problem and to improve the automated design of functional proteins are discussed. Copyright 1998 Academic Press

15/7/16 (Item 16 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09463100 98199336
Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches.
Cho SJ; Zheng W; Tropsha A
Laboratory for Molecular Modeling, School of Pharmacy, University of North Carolina, Chapel Hill 27599, USA.
Journal of chemical information and computer sciences (UNITED STATES) Mar-Apr 1998, 38 (2) p259-68, ISSN 0095-2338 Journal Code: HNT
Contract/Grant No.: MH 40537, MH, NIMH; HD03310, HD, NICHD; MH33127, MH, NIMH
Languages: ENGLISH
Document type: JOURNAL ARTICLE
We have developed a novel strategy for rational design of targeted peptide libraries. The goal of this method is to select a subset of natural amino acids that are most likely to be present in active peptides for the synthesis of library. Two different protocols are employed where chemical structures of peptides are described either by topological indices or by a combination of physicochemical descriptors for individual amino acids. The selection of a peptide as a candidate for the targeted library is based either on its chemical similarity to a biologically active probe or on its biological activity predicted from a preconstructed quantitative structure-activity (QSAR) equation. The optimization of the library is achieved by means of genetic algorithms (GA). This method was tested by rational design of the library with bradykinin-potentiating activity. Twenty-eight bradykinin-potentiating pentapeptides were used as a training set for the development of a QSAR equation, and, alternatively, two active pentapeptides, VEWAK and VKWAP, were used as probe molecules. In each case, the frequency distribution of amino acids in the top 100 peptides suggested by the method resembles the frequency distribution of amino acids found in the active peptides. The results obtained after GA optimization also compared favorably with those obtained by the exhaustive analysis of all possible 3.2 million pentapeptides.

15/7/17 (Item 17 from file: 155)

DIALOG(R)File 155:MEDLINE(R)

09425390 98136461
A genetic algorithm for multiple molecular sequence alignment.
Zhang C; Wong AK
Department of Systems Design Engineering, University of Waterloo, Ontario, Canada. czhang@watnow.uwaterloo.ca
Computer applications in the biosciences (ENGLAND) Dec 1997, 13 (6) p565-81, ISSN 0266-7061 Journal Code: CAB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
MOTIVATION: Multiple molecular sequence alignment is among the most important and most challenging tasks in computational biology. The currently used alignment techniques are characterized by great computational complexity, which prevents their wider use. This research is aimed at developing a new technique for efficient multiple sequence alignment. APPROACH: The new method is based on genetic algorithms . Genetic algorithms are stochastic approaches for efficient and robust searching. By converting biomolecular sequence alignment into a problem of searching for optimal or near-optimal points in an 'alignment space', a genetic algorithm can be used to find good alignments very efficiently. RESULTS: Experiments on real data sets have shown that the average computing time of this technique may be two or three orders lower than that of a technique based on pairwise dynamic programming, while the alignment qualities are very similar. AVAILABILITY: A C program on UNIX has been written to implement the technique. It is available on request from the authors.

15/7/18 (Item 18 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09424273 98114032
Strategies for identifying and predicting islet autoantigen T-cell epitopes in insulin-dependent diabetes mellitus.
Honeyman MC; Brusic V; Harrison LC
Autoimmunity and Transplantation Division, The Walter and Eliza Hall Institute, Royal Melbourne Hospital, Victoria, Australia. honeyman@wehi.edu.au
Annals of medicine (ENGLAND) Oct 1997, 29 (5) p401-4, ISSN 0785-3890 Journal Code: AMD
Languages: ENGLISH
Document type: JOURNAL ARTICLE
T cells recognize peptide epitopes bound to major histocompatibility complex molecules. Human T-cell epitopes have diagnostic and therapeutic applications in autoimmune diseases. However, their accurate definition within an autoantigen by T-cell bioassay, usually proliferation, involves many costly peptides and a large amount of blood. We have therefore developed a strategy to predict T-cell epitopes and applied it to tyrosine phosphatase IA-2, an autoantigen in IDDM, and HLA-DR4(*0401). First, the binding of synthetic overlapping peptides encompassing IA-2 was measured directly to purified DR4. Secondly, a large amount of HLA-DR4 binding data were analysed by alignment using a genetic algorithm and were used to train an artificial neural network to predict the affinity of binding. This bioinformatic prediction method was then validated experimentally and used

to predict DR4 binding peptides in IA-2. The binding set encompassed 85% of experimentally determined T-cell epitopes. Both the experimental and bioinformatic methods had high negative predictive values, 92% and 95%, indicating that this strategy of combining experimental results with computer modelling should lead to a significant reduction in the amount of blood and the number of peptides required to define T-cell epitopes in humans.

15/7/19 (Item 19 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09415115 98108003
Scoring functions for computational algorithms applicable to the design of spiked oligonucleotides.
Jensen LJ; Andersen KV; Svendsen A; Kretzschmar T
Department of Enzyme Design, Novo Nordisk A/S, DK-2880 Bagsvaerd, Denmark.
Nucleic acids research (ENGLAND) Feb 1 1998, 26 (3) p697-702, ISSN 0305-1048 Journal Code: O8L
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Protein engineering by inserting stretches of random DNA sequences into target genes in combination with adequate screening or selection methods is a versatile technique to elucidate and improve protein functions. Established compounds for generating semi-random DNA sequences are spiked oligonucleotides which are synthesised by interspersing wild type (wt) nucleotides of the target sequence with certain amounts of other nucleotides. Directed spiking strategies reduce the complexity of a library to a manageable format compared with completely random libraries. Computational algorithms render feasible the calculation of appropriate nucleotide mixtures to encode specified amino acid subpopulations. The crucial element in the ranking of spiked codons generated during an iterative algorithm is the scoring function. In this report three scoring functions are analysed: the sum-of-square-differences function s, a modified cubic function c, and a scoring function m derived from maximum likelihood considerations. The impact of these scoring functions on calculated amino acid distributions is demonstrated by an example of mutagenising a domain surrounding the active site serine of subtilisin-like proteases. At default weight settings of one for each amino acid, the new scoring function m is superior to functions s and c in finding matches to a given amino acid population.

15/7/20 (Item 20 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09354018 98051950
Computer simulations of prebiotic evolution.
Abkevich VI; Gutin AM; Shakhnovich EI
Harvard University, Department of Chemistry, Cambridge, MA 02138, USA.
Pacific Symposium on Biocomputing (SINGAPORE) 1997, p27-38,
Journal Code: CWQ
Languages: ENGLISH
Document type: JOURNAL ARTICLE

This paper is a review of our previous work on the field of possible ways of prebiotic evolution . We propose an algorithm providing sequences of model proteins with rapid folding into a given native conformation. Thermodynamical analysis shows that the increase in speed is matched by an increase in stability: the evolved sequences are much more stable in their native conformation than the initial random sequence. We discuss a possible origin of the first biopolymers, having stable unique structure. We suggest that at the prebiotic stage of evolution, long organic polymers had to be compact in order to avoid hydrolysis and had to be soluble and thus must not be exceedingly hydrophobic. We present an algorithm that generates such sequences of model proteins . The evolved sequences turn out to have a stable unique structure, into which they quickly fold. This result illustrates the idea that the unique three-dimensional native structure of first biopolymers could have evolved as a side effect of a nonspecific physico-chemical factors acting at the prebiotic stage of evolution.

15/7/21 (Item 21 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09251426 97446349
Structural analysis of a necrogenic strain of cucumber mosaic cucumovirus satellite RNA in planta.
Rodriguez-Alvarado G; Roossinck MJ
The S. R. Noble Foundation, Inc., Ardmore, Oklahoma, 73402, USA.
Virology (UNITED STATES) Sep 15 1997, 236 (1) p155-66, ISSN 0042-6822 Journal Code: XEA
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Structural studies of plant viral RNA molecules have been based on in vitro chemical and enzymatic modification. That approach, along with mutational analysis, has proven valuable in predicting structural models for some plant viruses such as tobacco mosaic tobamovirus and brome mosaic bromovirus. However, in planta conditions may be dramatically different from those found in vitro. In this study we analyzed the structure of cucumber mosaic cucumovirus satellite RNA (sat RNA) strain D4 in vivo and compared it to the structures found in vitro and in purified virions. Following a methodology developed to determine the structure of 18S rRNA within intact plant tissues, different patterns of adenosine and cytosine modification were found for D4-sat RNA molecules in vivo, in vitro, and in virions. This chemical probing procedure identifies adenosine and cytosine residues located in unpaired regions of the RNA molecules. Methylation data, a genetic algorithm in the STAR RNA folding program, and sequence alignment comparisons of 78 satellite CMV RNA sequences were used to identify several helical regions located at the 5' and 3' ends of the RNA molecule. Data from previous mutational and sequence comparison studies between satellite RNA strains inducing necrosis in tomato plants and those strains not inducing necrosis allowed us to identify one helix and two tetraloop regions correlating with the necrogenicity syndrome.

15/7/22 (Item 22 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09249959 97409488
A comparison of heuristic search algorithms for molecular docking.
Westhead DR; Clark DE; Murray CW
Proteus Molecular Design Ltd., Macclesfield, Cheshire, U.K.
Journal of computer-aided molecular design (NETHERLANDS) May 1997, 11 (3) p209-28, ISSN 0920-654X Journal Code: JCB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
This paper describes the implementation and comparison of four heuristic search algorithms (genetic algorithm , evolutionary programming, simulated annealing and tabu search) and a random search procedure for flexible molecular docking. To our knowledge, this is the first application of the tabu search algorithm in this area. The algorithms are compared using a recently described fast molecular recognition potential function and a diverse set of five protein-ligand systems. Statistical analysis of the results indicates that overall the genetic algorithm performs best in terms of the median energy of the solutions located. However, tabu search shows a better performance in terms of locating solutions close to the crystallographic ligand conformation. These results suggest that a hybrid search algorithm may give superior results to any of the algorithms alone.

15/7/23 (Item 23 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09106353 97206010
Optimizing doped libraries by using genetic algorithms.
Tomandl D; Schober A; Schwienhorst A
Department of Molecular Evolution Biology, Institute for Molecular Biotechnology, Jena, Germany.
Journal of computer-aided molecular design (NETHERLANDS) Jan 1997, 11 (1) p29-38, ISSN 0920-654X Journal Code: JCB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The insertion of random sequences into protein-encoding genes in combination with biological selection techniques has become a valuable tool in the design of molecules that have useful and possibly novel properties. By employing highly effective screening protocols, a functional and unique structure that had not been anticipated can be distinguished among a huge collection of inactive molecules that together represent all possible amino acid combinations. This technique is severely limited by its restriction to a library of manageable size. One approach for limiting the size of a mutant library relies on 'doping schemes', where subsets of amino acids are generated that reveal only certain combinations of amino acids in a protein sequence. Three mononucleotide mixtures for each codon concerned must be designed, such that the resulting codons that are assembled during chemical gene synthesis represent the desired amino acid mixture on the level of the translated protein. In this paper we present a doping algorithm that "reverse translates' a desired mixture of certain amino acids into three mixtures of mononucleotides. The algorithm is designed to optimally bias these mixtures towards the codons of choice. This approach combines a genetic algorithm with local optimization strategies based on the downhill simplex method. Disparate relative representations of all amino acids (and stop codons) within a target set can be generated. Optional weighing factors are employed to emphasize the frequencies of certain amino

acids and their codon usage, and to compensate for reaction rates of different mononucleotide building blocks (synthons) during chemical DNA synthesis. The effect of statistical errors that accompany an experimental realization of calculated nucleotide mixtures on the generated mixtures of amino acids is simulated. These simulations show that the robustness of different optima with respect to small deviations from calculated values depends on their concomitant fitness. Furthermore, the calculations probe the fitness landscape locally and allow a preliminary assessment of its structure.

15/7/24 (Item 24 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

09091163 97220237
Relation between amino acid composition and cellular location of proteins.
Cedano J; Aloy P; Perez-Pons JA; Querol E
Departament de Bioquimica i Biologia Molecular, Universitat Autonoma de Barcelona, Bellaterra, Spain.
Journal of molecular biology (ENGLAND) Feb 28 1997, 266 (3) p594-600, ISSN 0022-2836 Journal Code: J6V
Languages: ENGLISH
Document type: JOURNAL ARTICLE
A correlation analysis of the amino acid composition and the cellular location of a protein is presented. The statistical analysis discriminates among the following five protein classes: integral membrane proteins, anchored membrane proteins, extracellular proteins, intracellular proteins and nuclear proteins. This segregation into protein classes related to their location can help researchers to design experimental work for testing hypotheses in order to find out the functionality of a reading frame in search of function. A program (ProtLock) to predict the cellular location of a protein has been designed .

15/7/25 (Item 25 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

08911849 97020964
New approaches in molecular structure prediction.
Bohm G
Institut fur Biotechnologie, Martin-Luther-Universitat Halle-Wittenberg, Germany.
Biophysical chemistry (NETHERLANDS) Mar 7 1996, 59 (1-2) p1-32, ISSN 0301-4622 Journal Code: A5T
Languages: ENGLISH
Document type: JOURNAL ARTICLE; REVIEW; REVIEW, ACADEMIC
In the past years, much effort has been put on the development of new methodologies and algorithms for the prediction of protein secondary and tertiary structures from (sequence) data; this is reviewed in detail. New approaches for these predictions such as neural network methods, genetic algorithms , machine learning, and graph theoretical methods are discussed. Secondary structure prediction algorithms were improved mostly by considering families of related proteins; however, for the reliable tertiary structure modeling of proteins, knowledge-based techniques are

still preferred. Methods and examples with more or less successful results are described. Also, programs and parameterizations for energy minimisations, molecular dynamics, and electrostatic interactions have been improved, especially with respect to their former limits of applicability. Other topics discussed in this review include the use of traditional and on-line databases, the docking problem and surface properties of biomolecules, packing of protein cores, de novo design and protein engineering, prediction of membrane protein structures, the verification and reliability of model structures, and progress made with currently available software and computer hardware. In summary, the prediction of the structure, function, and other properties of a protein is still possible only within limits, but these limits continue to be moved. (364 Refs.)

15/7/26 (Item 26 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

08855595 97001877
Exploring the energy landscapes of molecular recognition by a genetic algorithm : analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes.
Verkhivker GM; Rejto PA; Gehlhaar DK; Freer ST
Agouron Pharmaceuticals, Inc., San Diego, California 92121, USA.
Proteins (UNITED STATES) Jul 1996, 25 (3) p342-53, ISSN 0887-3585
Journal Code: PTS
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Energy landscapes of molecular recognition are explored by performing "semi-rigid" docking of FK-506 and rapamycin with the Fukisawa binding protein (FKBP-12), and flexible docking simulations of the Ro-31-8959 and AG-1284 inhibitors with HIV-1 protease by a genetic algorithm . The requirements of a molecular recognition model to meet thermodynamic and kinetic criteria of ligand-protein docking simultaneously are investigated using a family of simple molecular recognition energy functions. The critical factor that determines the success rate in predicting the structure of ligand-protein complexes is found to be the roughness of the binding energy landscape, in accordance with a minimal frustration principle. The results suggest that further progress in structure prediction of ligand- protein complexes can be achieved by designing molecular recognition energy functions that generate binding landscapes with reduced frustration.

15/7/27 (Item 27 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

08619791 96149395
How the first biopolymers could have evolved.
Abkevich VI; Gutin AM; Shakhnovich EI
Harvard University, Department of Chemistry, Cambridge, MA 02138, USA.
Proceedings of the National Academy of Sciences of the United States of America (UNITED STATES) Jan 23 1996, 93 (2) p839-44, ISSN 0027-8424
Journal Code: PV3
Languages: ENGLISH
Document type: JOURNAL ARTICLE

In this work, we discuss a possible origin of the first biopolymers with stable unique structures. We suggest that at the prebiotic stage of evolution, long organic polymers had to be compact to avoid hydrolysis and had to be soluble and thus must not be exceedingly hydrophobic. We present an algorithm that generates such sequences for model proteins . The evolved sequences turn out to have a stable unique structure, into which they quickly fold. This result illustrates the idea that the unique three-dimensional native structures of first biopolymers could have evolved as a side effect of nonspecific physicochemical factors acting at the prebiotic stage of evolution.

15/7/28 (Item 28 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

08471347 96117651
De novo design of the hydrophobic cores of proteins.
Desjarlais JR; Handel TM
Department of Molecular and Cell Biology, University of California at Berkeley 94720, USA.
Protein science (UNITED STATES) Oct 1995, 4 (10) p2006-18, ISSN 0961-8368 Journal Code: BNW
Languages: ENGLISH
Document type: JOURNAL ARTICLE
We have developed and experimentally tested a novel computational approach for the de novo design of hydrophobic cores. A pair of computer programs has been written, the first of which creates a "custom" rotamer library for potential hydrophobic residues, based on the backbone structure of the protein of interest. The second program uses a genetic algorithm to globally optimize for a low energy core sequence and structure, using the custom rotamer library as input. Success of the programs in predicting the sequences of native proteins indicates that they should be effective tools for protein design . Using these programs, we have designed and engineered several variants of the phage 434 cro protein, containing five, seven, or eight sequence changes in the hydrophobic core. As controls, we have produced a variant consisting of a randomly generated core with six sequence changes but equal volume relative to the native core and a variant with a "minimalist" core containing predominantly leucine residues. Two of the designs, including one with eight core sequence changes, have thermal stabilities comparable to the native protein , whereas the third design and the minimalist protein are significantly destabilized. The randomly designed control is completely unfolded under equivalent conditions. These results suggest that rational de novo design of hydrophobic cores is feasible, and stress the importance of specific packing interactions for the stability of proteins. A surprising aspect of the results is that all of the variants display highly cooperative thermal denaturation curves and reasonably dispersed NMR spectra. This suggests that the non-core residues of a protein play a significant role in determining the uniqueness of the folded structure.

15/7/29 (Item 29 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

08465819 96133308

A genetic algorithm that seeks native states of peptides and proteins.

Sun S

Structural Biochemistry Program, Frederick Biomedical Supercomputing Center, National Cancer Institute, Frederick Cancer Research and Development Center, Maryland 21702, USA.

Biophysical journal (UNITED STATES) Aug 1995, 69 (2) p340-55, ISSN 0006-3495 Journal Code: A5S

Languages: ENGLISH

Document type: JOURNAL ARTICLE

We describe a computer algorithm to predict native structures of proteins and peptides from their primary sequences, their known native radii of gyration, and their known disulfide bonding patterns, starting from random conformations. Proteins are represented as simplified real-space main chains with single-bead side chains. Nonlocal interactions are taken from structural database-derived statistical potentials, as in an earlier treatment. Local interactions are taken from simulations of (phi, psi) energy surfaces for each amino acid generated using the Biosym Discover program. Conformational searching is done by a genetic algorithm -based method. Reasonable structures are obtained for melittin (a 26-mer), avian pancreatic polypeptide inhibitor (a 36-mer), crambin (a 46-mer), apamin (an 18-mer), tachyplesin (a 17-mer), C-peptide of ribonuclease A (a 13-mer), and four different designed helical peptides . A hydrogen bond interaction was tested and found to be generally unnecessary for helical peptides, but it helps fold some sheet regions in these structures. For the few longer chains we tested, the method appears not to converge. In those cases, it appears to recover native-like secondary structures, but gets incorrect tertiary folds.

15/7/30 (Item 30 from file: 155)

DIALOG(R)File 155:MEDLINE(R)

08396445 95396117

An APL-programmed genetic algorithm for the prediction of RNA secondary structure.

van Batenburg FH; Gultyaev AP; Pleij CW

Institute for Theoretical Biology, Leiden, The Netherlands.

Journal of theoretical biology (ENGLAND) Jun 7 1995, 174 (3) p269-80, ISSN 0022-5193 Journal Code: K8N

Languages: ENGLISH

Document type: JOURNAL ARTICLE

The possibilities of using a genetic algorithm for the prediction of RNA secondary structure were investigated. The algorithm, using the procedure of stepwise selection of the most fit structures (similarly to natural evolution), allows different models of fitness or driving forces determining RNA structure to be easily introduced. This can be used for simulation of the RNA folding process and for the investigation of possible folding pathways. Such an algorithm needs several modifications before it can predict RNA secondary structures. After modification , a fair number of correct stems are predicted, even when using computationally quick, but very crude, fitness criteria such as stem length and stacking energy, including elements of tertiary structure (pseudoknots). The fact that genetic algorithm simulation includes both stem formations and stem disruption allows one to observe intermediate structures that may be used in combination with phylogenetic or experimental research.

15/7/31 (Item 31 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

08389380 95192051
A genetic algorithm based molecular modeling technique for RNA stem-loop structures.
Ogata H; Akiyama Y; Kanehisa M
Institute for Chemical Research, Kyoto University, Japan.
Nucleic acids research (ENGLAND) Feb 11 1995, 23 (3) p419-26, ISSN 0305-1048 Journal Code: O8L
Languages: ENGLISH
Document type: JOURNAL ARTICLE
A new modeling technique for arriving at the three dimensional (3-D) structure of an RNA stem-loop has been developed based on a conformational search by a genetic algorithm and the following refinement by energy minimization. The genetic algorithm simultaneously optimizes a population of conformations in the predefined conformational space and generates 3-D models of RNA . The fitness function to be optimized by the algorithm has been defined to reflect the satisfaction of known conformational constraints. In addition to a term for distance constraints, the fitness function contains a term to constrain each local conformation near to a prepared template conformation. The technique has been applied to the two loops of tRNA, the anticodon loop and the T-loop, and has found good models with small root mean square deviations from the crystal structure. Slightly different models have also been found for the anticodon loop. The analysis of a collection of alternative models obtained has revealed statistical features of local variations at each base position.

15/7/32 (Item 32 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

07972020 94322918
Rapid evolution of a protein in vitro by DNA shuffling [see comments]
Stemmer WP
Affymax Research Institute, Palo Alto, California 94304.
Nature (ENGLAND) Aug 4 1994, 370 (6488) p389-91, ISSN 0028-0836 Journal Code: NSC
Comment in Nature 1994 Aug 4;370(6488):324-5
Languages: ENGLISH
Document type: JOURNAL ARTICLE
DNA shuffling is a method for in vitro homologous recombination of pools of selected mutant genes by random fragmentation and polymerase chain reaction (PCR) reassembly. Computer simulations called genetic algorithms have demonstrated the importance of iterative homologous recombination for sequence evolution. Oligonucleotide cassette mutagenesis and error-prone PCR are not combinatorial and thus are limited in searching sequence space. We have tested mutagenic DNA shuffling for molecular evolution in a beta-lactamase model system. Three cycles of shuffling and two cycles of backcrossing with wild-type DNA , to eliminate non-essential mutations, were each followed by selection on increasing concentrations of the antibiotic cefotaxime. We report here that selected mutants had a minimum inhibitory concentration of 640 micrograms ml-1, a 32,000-fold

increase and 64-fold greater than any published TEM-1 derived enzyme. Cassette mutagenesis and error-prone PCR resulted in only a 16-fold increase.

15/7/33 (Item 33 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

07939957 94272340
De novo protein design using pairwise potentials and a genetic algorithm.
Jones DT
Department of Biochemistry and Molecular Biology, University College, London United Kingdom.
Protein science (UNITED STATES) Apr 1994, 3 (4) p567-74, ISSN 0961-8368 Journal Code: BNW
Languages: ENGLISH
Document type: JOURNAL ARTICLE
One of the major goals of molecular biology is to understand how protein chains fold into a unique 3-dimensional structure. Given this knowledge, perhaps the most exciting prospect will be the possibility of designing new proteins to perform designated tasks, an application that could prove to be of great importance in medicine and biotechnology. It is possible that effective protein design may be achieved without the requirement for a full understanding of the protein folding process. In this paper a simple method is described for designing an amino acid sequence to fit a given 3-dimensional structure. The compatibility of a designed sequence with a given fold is assessed by means of a set of statistically determined potentials (including interresidue pairwise and solvation terms), which have been previously applied to the problem of protein fold recognition. In order to generate sequences that best fit the fold, a genetic algorithm is used, whereby the sequence is optimized by a stochastic search in the style of natural selection.

15/7/34 (Item 34 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

07740494 94192007
SIMD parallelization of the WORDUP algorithm for detecting statistically significant patterns in DNA sequences.
Liuni S; Prunella N; Pesole G; D'Orazio T; Stella E; Distante A
CSMME-CNR, Bari, Italy.
Computer applications in the biosciences (ENGLAND) Dec 1993, 9 (6) p701-7, ISSN 0266-7061 Journal Code: CAB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The development of new techniques in sequencing nuclei acids has produced a great amount of sequence data and has led to the discovery of new relationships. In this paper, we study a method for parallelizing the algorithm WORDUP, which detects the presence of statistically significant patterns in DNA sequences. WORDUP implements an efficient method to identify the presence of statistically significant oligomers in a non-homologous group of sequences. It is based on a modified version of the Boyer-Moore algorithm, which is one of the fastest algorithms for string

matching available in the literature. The aim of the parallel version of WORDUP presented here is to speed up the computational time and allow the analysis of a greater set of longer nucleotide sequences, which is usually impractical with sequential algorithms.

15/7/35 (Item 35 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

07337420 90219075
The TEACL method of DNA-DNA hybridization: technical considerations.
Powell JR; Caccone A
Department of Biology, Yale University, New Haven, Connecticut 06511.
Journal of molecular evolution (UNITED STATES) Mar 1990, 30 (3) p267-72, ISSN 0022-2844 Journal Code: J76
Languages: ENGLISH
Document type: JOURNAL ARTICLE
This paper emanated from a conference concerning the value, accuracy, and technical considerations of DNA-DNA hybridization for evolutionary studies. Our laboratory has been performing the so-called TEACL (tetraethylammonium chloride) method, and we have amassed sufficient data to indicate that this method is very powerful if performed properly with correct analyses. Here we address five technical considerations: (1) We present empirical data that size correction for tracer length is legitimate and accurate. (2) We show that the error of delta Tm measurement does not significantly increase with increasing distance up to at least 10 degrees C. (3) The error distribution for delta Tm does not deviate from the expected normal distribution indicating parametric statistics are probably legitimate for analyses. (4) Using a known phylogeny we examined the resolving power of the technique by showing that at least five taxa can be correctly placed in phylogenies with a maximum delta Tm of 2.5 degrees C. (5) To date, all our data sets based on DNA-DNA hybridization are very robust with respect to analytical procedures in that every algorithm used on the data sets has yielded identical trees with nearly identical branch lengths. Nevertheless, we point out that theoretical analyses of distance data (as generated by DNA -DNA hybridization) are lacking, especially with regard to tests of the molecular clock hypothesis.

15/7/36 (Item 36 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

07330307 93108437
Mathematical characterization of Chaos Game Representation. New algorithms for nucleotide sequence analysis.
Dutta C; Das J
Biophysics Division, Indian Institute of Chemical Biology, Calcutta.
Journal of molecular biology (ENGLAND) Dec 5 1992, 228 (3) p715-9, ISSN 0022-2836 Journal Code: J6V
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Chaos Game Representation (CGR) can recognize patterns in the nucleotide sequences, obtained from databases, of a class of genes using the techniques of fractal structures and by considering DNA sequences as strings composed of four units, G, A, T and C. Such recognition of patterns

relies only on visual identification and no mathematical characterization of CGR is known. The present report describes two algorithms that can predict the presence or absence of a stretch of nucleotides in any gene family. The first algorithm can be used to generate DNA sequences represented by any point in the CGR. The second algorithm can simulate known CGR patterns for different gene families by setting the probabilities of occurrence of different di- or trinucleotides by a trial and error process using some guidelines and approximate rules-of-thumb. The validity of the second algorithm has been tested by simulating sequences that can mimic the CGRs of vertebrate non-oncogenes, proto-oncogenes and oncogenes. These algorithms can provide a mathematical basis of the CGR patterns obtained using nucleotide sequences from databases.

15/7/37 (Item 37 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

07077765 92339042
The rapid generation of mutation data matrices from protein sequences.
Jones DT; Taylor WR; Thornton JM
Department of Biochemistry and Molecular Biology, University College, London, UK.
Computer applications in the biosciences (ENGLAND) Jun 1992, 8 (3) p275-82, ISSN 0266-7061 Journal Code: CAB
Languages: ENGLISH
Document type: JOURNAL ARTICLE
An efficient means for generating mutation data matrices from large numbers of protein sequences is presented here. By means of an approximate peptide-based sequence comparison algorithm, the set sequences are clustered at the 85% identity level. The closest relating pairs of sequences are aligned, and observed amino acid exchanges tallied in a matrix. The raw mutation frequency matrix is processed in a similar way to that described by Dayhoff et al. (1978), and so the resulting matrices may be easily used in current sequence analysis applications, in place of the standard mutation data matrices, which have not been updated for 13 years. The method is fast enough to process the entire SWISS-PROT databank in 20 h on a Sun SPARCstation 1, and is fast enough to generate a matrix from a specific family or class of proteins in minutes. Differences observed between our 250 PAM mutation data matrix and the matrix calculated by Dayhoff et al. are briefly discussed.

15/7/38 (Item 38 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

06621960 90190399
Rapid and sensitive sequence comparison with FASTP and FASTA.
Pearson WR
Methods in enzymology (UNITED STATES) 1990, 183 p63-98, ISSN 0076-6879 Journal Code: MVA
Languages: ENGLISH
Document type: JOURNAL ARTICLE
The FASTA program can search the NBRF protein sequence library (2.5 million residues) in less than 20 min on an IBM-PC microcomputer and

unambiguously detect proteins that shared a common ancestor billions of years in the past. FASTA is both fast and selective because it initially considers only amino acid identities. Its sensitivity is increased not only by using the PAM250 matrix to score and rescore regions with large numbers of identities but also by joining initial regions. The results of searches with FASTA compare favorably with results using NWS-based programs that are 100 times slower. FASTA is slightly less sensitive but considerably more selective. It is not clear that NWS-based programs would be more successful in finding distantly related members of the G-protein-coupled receptor family. The joining step by FASTA to calculate the initn score is especially useful for sequences that share regions of sequence similarity that are separated by variable-length loops. FASTP and FASTA were designed to identify protein sequences that have descended from a common ancestor, and they have proved very useful for this task. In many cases, a FASTA sequence search will result in a list of high scoring library sequences that are homologous to the query sequence, or the search will result in a list of sequences with similarity scores that cannot be distinguished from the bulk of the library. In either case, the question of whether there are sequences in the library that are clearly related to the query sequence has been answered unambiguously. Unfortunately, the results often will not be so clear-cut, and careful analysis of similarity scores, statistical significance, the actual aligned residues, and the biological context are required. In the course of analyzing the G-protein-coupled receptor family, several proteins were found that, because of a high initn score and a low init1 score that increased almost 2-fold with optimization, appeared to be members of this family which were not previously recognized. RDF2 analysis showed borderline z values, and only a careful examination of the sequence alignments that focused on the conserved residues provided convincing evidence that the high scores were fortuitous. As sequence comparison methods become more powerful by becoming more sensitive, they become more likely to mislead, and even greater care is required.

15/7/39 (Item 39 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

06013481 85142155
Rapid and sensitive protein similarity searches.
Lipman DJ; Pearson WR
Science (UNITED STATES) Mar 22 1985, 227 (4693) p1435-41, ISSN 0036-8075 Journal Code: UJ7
Contract/Grant No.: SO7-RR05431, RR, NCRR
Languages: ENGLISH
Document type: JOURNAL ARTICLE
An algorithm was developed which facilitates the search for similarities between newly determined amino acid sequences and sequences already available in databases. Because of the algorithm's efficiency on many microcomputers, sensitive protein database searches may now become a routine procedure for molecular biologists. The method efficiently identifies regions of similar sequence and then scores the aligned identical and differing residues in those regions by means of an amino acid replacability matrix. This matrix increases sensitivity by giving high scores to those amino acid replacements which occur frequently in evolution . The algorithm has been implemented in a computer program designed to search protein databases very rapidly. For example, comparison of a 200-amino-acid sequence to the 500,000 residues in the

National Biomedical Research Foundation library would take less than 2 minutes on a minicomputer, and less than 10 minutes on a microcomputer (IBM PC).

15/7/40 (Item 40 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

05349889 88302231
[Evolution of RNA-dependent RNA-polymerases from positive RNA viruses: comparison of phylogenetic trees constructed by different methods]
Evoliutsiia RNK-zavisimykh RNK-polimeraz pozitivnykh ribovirusov: sravnenie filogeneticheskikh derev'ev, postroennykh raznymi metodami.
Kunin EV; Chumakov KM; Iushmanov SV; Gorbalenia AE
Molekuliarnaia genetika, mikrobiologiia i virusologiia (USSR) Mar 1988, (3) p16-9, ISSN 0208-0613 Journal Code: NMJ
Languages: RUSSIAN Summary Languages: ENGLISH
Document type: JOURNAL ARTICLE ; English Abstract
Presumptive phylogenetic trees of evolutionary conserved fragments of RNA-dependent RNA polymerases of 26 positive strand RNA viruses were generated using a simple clustering procedure or a novel approach based on the so-called maximal topologic similarity principle. The latter methodology involves a quantitative measure of the degree of correspondence between the topology of generated trees and structure of the initial distance matrix. The algorithm for tree construction based on the maximal topologic similarity principle does not include the assumption of evolutionary rate constancy, as opposed to the clustering procedure. Nevertheless, it is demonstrated that the trees generated by the two methods are topologically similar, indicating that no drastic change of evolutionary rate had occurred in evolution of the positive strand RNA virus RNA polymerases. This in turn suggests that RNA-dependent RNA polymerases (or at least their evolutionary conserved core domains used for construction of the phylogenetic trees) are principally functionally equivalent in all positive strand RNA viruses.

15/7/41 (Item 41 from file: 155)
DIALOG(R)File 155:MEDLINE(R)

05180969 87175673
Construction of multilocus genetic linkage maps in humans.
Lander ES; Green P
Proceedings of the National Academy of Sciences of the United States of America (UNITED STATES) Apr 1987, 84 (8) p2363-7, ISSN 0027-8424
Journal Code: PV3
Languages: ENGLISH
Document type: JOURNAL ARTICLE
Human genetic linkage maps are most accurately constructed by using information from many loci simultaneously. Traditional methods for such multilocus linkage analysis are computationally prohibitive in general, even with supercomputers. The problem has acquired practical importance because of the current international collaboration aimed at constructing a complete human linkage map of DNA markers through the study of three-generation pedigrees. We describe here several alternative algorithms for constructing human linkage maps given a specified gene order. One

method allows maximum-likelihood multilocus linkage maps for dozens of DNA markers in such three generation pedigrees to be constructed in minutes.

15/7/42 (Item 1 from file: 2)
DIALOG(R)File 2:INSPEC

6813111 INSPEC Abstract Number: A2001-04-8715B-013, C2001-02-7330-296
Title: Prediction of protein structures using a Hopfield network
Author(s): Scott, L.P.B.; Chahine, J.; Ruggiero, J.R.
Conference Title: Proceedings. Vol.1. Sixth Brazilian Symposium on Neural Networks p.284
Editor(s): Ribeiro, C.H.C.; Franca, F.M.G.
Publisher: IEEE Comput. Soc, Los Alamitos, CA, USA
Publication Date: 2000 Country of Publication: USA xii+296 pp.
ISBN: 0 7695 0856 1 Material Identity Number: XX-2000-02632
U.S. Copyright Clearance Center Code: 1522 4899/2000/$10.00
Conference Title: Proceedings Sixth Brazilian Symposium on Neural Networks
Conference Sponsor: Brazilian Comput. Soc. (SBC); Special Interest Group of the Int. Neural Networks Soc. Brazil (SIG/INNS/BR)
Conference Date: 22-25 Nov. 2000 Conference Location: Rio de Janeiro, RJ, Brazil
Language: English Document Type: Conference Paper (PA)
Treatment: Practical (P)
Abstract: Summary form only given. Under proper conditions, a globular protein adopts a unique 3D structure that is encoded in an amino acid sequence. The theoretical prediction of this structure, and the pathways followed during the folding process, are an important problem in structural molecular biology. Several works have explored the application of genetic algorithms and neural networks to the determination of the protein structure. There are several techniques of computational simulation that can be used to study structure of proteins; methods of Monte Carlo, simulated annealing, genetic algorithms and neural networks. This work discusses the possibilities to use neural networks in the study of macromolecule structures and presents a example of a Hopfield network to predict the structure of a protein and discusses the results and possible future works using neural networks and genetic algorithms to design new proteins and drugs. This paper used a Hopfield network to predict a primary sequence and the tertiary structure of the core of the cytochrome b/sub 562/. The neural network was implemented using the programming language C and the simulations were run on Silicon Graphics. (0 Refs)
Subfile: A C

15/7/43 (Item 2 from file: 2)
DIALOG(R)File 2:INSPEC

6778262 INSPEC Abstract Number: C2001-01-1340F-016
Title: DNA genetic algorithm for design of the generalized membership-type Takagi-Sugeno fuzzy control system
Author(s): Yongsheng Ding; Lihong Ren
Author Affiliation: Dept. of Autom., Dong Hua Univ., Shanghai, China
Conference Title: SMC 2000 Conference Proceedings. 2000 IEEE

International Conference on Systems, Man and Cybernetics. `Cybernetics Evolving to Systems, Humans, Organizations, and their Complex Interactions' (Cat. No.00CH37166) Part vol.5 p.3862-7 vol.5
Publisher: IEEE, Piscataway, NJ, USA
Publication Date: 2000 Country of Publication: USA 5 vol.3895 pp.
ISBN: 0 7803 6583 6 Material Identity Number: XX-2000-02510
U.S. Copyright Clearance Center Code: 0 7803 6583 6/2000/$10.00
Conference Title: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics
Conference Sponsor: Syst., Man and Cybern. Soc. IEEE
Conference Date: 8-11 Oct. 2000 Conference Location: Nashville, TN, USA
Language: English Document Type: Conference Paper (PA)
Treatment: Theoretical (T)
Abstract: We propose a new DNA-based genetic algorithm (DNA -GA) to optimize the design parameters of a generalized membership-type Takagi-Sugeno fuzzy controller (GTSFC). The GTSFC employs TS fuzzy rules with linear consequent, e/sup - ax+b (c)/-type input fuzzy sets containing almost arbitrary continuous input fuzzy sets, Zadeh fuzzy logic AND operation, and the widely-used centroid defuzzier. The GTSFC is proved to be a nonlinear PI controller with variable gains. The optimized design parameters are the input fuzzy sets and the linear consequent of the rules. The DNA-GA uses a DNA encoding method stemmed from the structure of the biological DNA to encode the design parameters of the GTSFC. The genetic operators of the method are based on the DNA genetic operations. The encoding method can significantly shorten the code length of DNA chromosomes and is suitable for complex knowledge representation. As a demonstration, we show how to implement the new method to optimize the design parameters of the GTSFC to control a nonlinear system. Computer simulation results indicate that the performance of the designed fuzzy controller is satisfactory. (23 Refs)
Subfile: C

15/7/44 (Item 3 from file: 2)
DIALOG(R)File 2:INSPEC

6629669 INSPEC Abstract Number: C2000-08-1180-011
Title: DNA genetic algorithms for design of fuzzy systems
Author(s): Lihong Ren; Yongsheng Ding; Shihuang Shao
Author Affiliation: Dept. of Autom., Donghua Univ., Shanghai, China
Conference Title: Ninth IEEE International Conference on Fuzzy Systems. FUZZ- IEEE 2000 (Cat. No.00CH37063) Part vol.2 p.1005-8 vol.2
Publisher: IEEE, Piscataway, NJ, USA
Publication Date: 2000 Country of Publication: USA 2 vol. xx+1080 pp.
ISBN: 0 7803 5877 5 Material Identity Number: XX-2000-00918
U.S. Copyright Clearance Center Code: 0 7803 5877 5/2000/$10.00
Conference Title: Ninth IEEE International Conference on Fuzzy Systems. FUZZ-IEEE 2000. Soft Computing in the Information Age
Conference Sponsor: IEEE Neural Networks Council; Texas A&M Univ.; Int. Fuzzy Syst. Assoc.(IFSA); Japan Soc. Fuzzy Theory & Syst.(SOFT)
Conference Date: 7-10 May 2000 Conference Location: San Antonio, TX, USA
Language: English Document Type: Conference Paper (PA)
Treatment: Applications (A); Theoretical (T)

Abstract: A new DNA genetic algorithm (DNA-GA) based on the mechanism of biological DNA and genetic information is proposed. The genetic operators of the DNA-GA are discussed. The DNA encoding method is suitable for the representation of complex knowledge. The DNA -GA is employed to design effective generalized fuzzy systems (GFS) for the modeling and control applications. GFS employ arbitrary fuzzy rules, e(- alpha /sub 1/x/sub 1/+ alpha /sub 2/)-type input fuzzy sets containing almost arbitrary continuous input fuzzy sets, arbitrary singleton output fuzzy sets, arbitrary fuzzy logic AND, and the generalized defuzzifier containing the widely-used centroid defuzzifier as a special case. The DNA-GA is used to select input variables and to tune the design parameters and membership functions of GFS. As such, the fuzzy rule sets of GFS can be obtained. The work in this paper provides a useful way for the design of fuzzy controllers and fuzzy models (14 Refs)

Subfile: C

15/7/45 (Item 4 from file: 2)
DIALOG(R)File 2:INSPEC

6506869 INSPEC Abstract Number: C2000-03-4290-011

Title: Virtual DNA simulator and protocol design by GA

Author(s): Nishikawa, A.; Hagiya, M.; Yamamura, M.

Author Affiliation: Dept. of Inf. Sci., Tokyo Univ., Japan

Conference Title: GECCO-99. Proceedings of the Genetic and Evolutionary Computation Conference. Joint Meeting of the Eighth International Conference on Genetic Algorithms (ICGA-99) and the Fourth Annual Genetic Programming Conference (GP-99) Part vol.2 p.1810-16 vol.2

Editor(s): Banzhaf, W.; Daida, J.; Eiben, A.E.; Garzon, M.H.; Honavar, V. ; Jakiela, M.; Smith, R.E.

Publisher: Morgan Kaufmann Publishers, San Francisco, CA, USA

Publication Date: 1999 Country of Publication: USA 2 vol. xvi+1876 pp.

ISBN: 1 55860 611 4 Material Identity Number: XX-2000-00240

Conference Title: Proceedings GECCO-99. Genetic and Evolutionary Computation Conference. Eighth International Conference on Genetic Algorithms (ICGA-99) and the Fourth Annual Gentic Programming Conference (GP-99)

Conference Date: 13-17 July 1999 Conference Location: Orlando, FL, USA

Language: English Document Type: Conference Paper (PA)

Treatment: Theoretical (T)

Abstract: Many algorithms and protocols for DNA computing have been proposed, but most of them remain mere proposals and their feasibility has not yet been verified. Even in cases when in vitro experiments are possible, it is desirable to verify the feasibility in advance. We developed a simulator to aid those who design algorithms and protocols for DNA computing. In this simulator, abstract sequences instead of real DNA sequences are used to represent molecules in order to increase efficiency of simulations. It consists of two main parts, one for finding reactions among existing molecules and generating new ones, and the other for numerically solving differential equations to calculate the concentration of each molecule. The two parts rely on each other. In particular, the former avoids a combinatorial explosion by setting a threshold on concentrations of molecules that can take part in reactions. Some simulation results are also presented: computation of Boolean

circuits, formation of DNA tiles and simulation of polymerase chain reaction (PCR). As for PCR, we also tried to find good protocols for PCR amplification using a genetic algorithm (GA). (10 Refs)
Subfile: C
Copyright 2000, IEE

15/7/46 (Item 5 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2001 Institution of Electrical Engineers. All rts. reserv.

6506849 INSPEC Abstract Number: C2000-03-7330-534
Title: Using evolutionary algorithms in the design of protein fingerprints
Author(s): Olsson, B.
Author Affiliation: Dept. of Comput. Sci., Skovde Univ., Sweden
Conference Title: GECCO-99. Proceedings of the Genetic and Evolutionary Computation Conference. Joint Meeting of the Eighth International Conference on Genetic Algorithms (ICGA-99) and the Fourth Annual Genetic Programming Conference (GP-99) Part vol.2 p.1636-42 vol.2
Editor(s): Banzhaf, W.; Daida, J.; Eiben, A.E.; Garzon, M.H.; Honavar, V.; Jakiela, M.; Smith, R.E.
Publisher: Morgan Kaufmann Publishers, San Francisco, CA, USA
Publication Date: 1999 Country of Publication: USA 2 vol. xvi+1876 pp.
ISBN: 1 55860 611 4 Material Identity Number: XX-2000-00240
Conference Title: Proceedings GECCO-99. Genetic and Evolutionary Computation Conference. Eighth International Conference on Genetic Algorithms (ICGA-99) and the Fourth Annual Gentic Programming Conference (GP-99)
Conference Date: 13-17 July 1999 Conference Location: Orlando, FL, USA
Language: English Document Type: Conference Paper (PA)
Treatment: Practical (P)
Abstract: This paper shows how evolutionary algorithms (EAs) are used as components in a system for design of protein fingerprints. The system is used for automated mining of data from protein sequence databases, with the purpose of deriving protein family fingerprints. The fingerprints are expressed as patterns, which can be used for recognition of sequences belonging to specific protein families. The system constructs candidate patterns by analyzing multiple sequence alignments, and selecting pattern elements corresponding to evolutionary conserved positions. Since most candidate patterns are too specific, we use stochastic search algorithms for generalization of the candidate patterns. In a previous version of the system a hill-climbing algorithm was used. We show how results can be substantially improved by using EAs for this task. We also compare a "standard" EA with a host-parasite EA, and show that it can significantly reduce the number of evaluations. (19 Refs)
Subfile: C
Copyright 2000, IEE

15/7/47 (Item 6 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2001 Institution of Electrical Engineers. All rts. reserv.

6459828 INSPEC Abstract Number: C2000-02-4290-002
Title: Virus-enhanced genetic algorithms inspired by DNA computing

Author(s): Mulawka, J.J.; Wasiewicz, P.; Pietak, K.
Author Affiliation: Warsaw Univ. of Technol., Poland
Conference Title: Foundations of Intelligent Systems. 11th International Symposium, ISMIS'99. Proceedings p.529-37
Editor(s): Ras, Z.W.; Skowron, A.
Publisher: Springer-Verlag, Berlin, Germany
Publication Date: 1999 Country of Publication: Germany xii+676 pp.
ISBN: 3 540 65965 X Material Identity Number: XX-1999-01669
Conference Title: Proceedings of ISMIS'99: 11th International Symposium on Methodologies for Intelligent Systems
Conference Sponsor: ICS PAS; Polish-Japanese Sch. Inf. Technol
Conference Date: 8-11 June 1999 Conference Location: Warsaw, Poland
Language: English Document Type: Conference Paper (PA)
Treatment: Theoretical (T)
Abstract: DNA computing is a new promising paradigm to develop an alternative generation of computers. Such approach is based on biochemical reactions using DNA strands which should be carefully designed . To this purpose a special DNA sequences design tool is required. The primary objective of this contribution is to present virus-enhanced genetic algorithms for global optimization to create a set of DNA strands. The main feature of the algorithms are mechanisms included specially for searching solution space of problems with complex bounds. Formulae, describing bounds of power of sequences' sets, which satisfy criteria and estimation functions are expressed. A computer program, called Mismatch, was implemented in C++ and runs on Windows NT platform. (13 Refs)
Subfile: C

15/7/48 (Item 7 from file: 2)
DIALOG(R)File 2:INSPEC

6338918 INSPEC Abstract Number: B1999-10-2230B-007, C1999-10-4290-010
Title: *Physical separation of DNA according to Royal Road fitness*
Author(s): Harlan Wood, D.; Junghuei Chen
Author Affiliation: Dept. of Comput. & Inf. Sci., Delaware Univ., Newark, DE, USA
Conference Title: Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406) Part Vol. 2 p.1011-16 Vol. 2
Publisher: IEEE, Piscataway, NJ, USA
Publication Date: 1999 Country of Publication: USA 3 vol. (xxxvii+2348) pp.
ISBN: 0 7803 5536 9 Material Identity Number: XX-1999-02118
U.S. Copyright Clearance Center Code: 0 7803 5536 9/99/$10.00
Conference Title: Proceedings of the 1999. Congress on Evolutionary Computation-CEC99
Conference Date: 6-9 July 1999 Conference Location: Washington, DC, USA
Language: English Document Type: Conference Paper (PA)
Treatment: Theoretical (T)
Abstract: We want to implement evolutionary computation using DNA, with trillions of candidate solutions being simultaneously evaluated for fitness. Unsurprisingly, the most difficult aspect is designing and implementing laboratory methods for physical separation of DNA strands according to "fitness". We propose a DNA strand design suited to the classical Royal Road Problem (E. van Nimwegen, et al.). We also propose

companion laboratory operations which would physically separate these DNA strands according to the Royal Road fitness criterion. (41 Refs)
Subfile: B C

15/7/49 (Item 8 from file: 2)
DIALOG(R)File 2:INSPEC

6338817 INSPEC Abstract Number: C1999-10-1180-013
Title: Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)
Part Vol. 1
Publisher: IEEE, Piscataway, NJ, USA
Publication Date: 1999 Country of Publication: USA 3 vol. (xxxvii+2348) pp.
ISBN: 0 7803 5536 9 Material Identity Number: XX-1999-02117
U.S. Copyright Clearance Center Code: 99/$10.00
Conference Title: Proceedings of the 1999. Congress on Evolutionary Computation-CEC99
Conference Date: 6-9 July 1999 Conference Location: Washington, DC, USA
Language: English Document Type: Conference Proceedings (CP)
Abstract: The following topics are dealt with: evolutionary computation and genetic algorithms ; multi-objective optimization; machine learning; time series; robotics; biomodeling and artificial life; engineering design ; constraint handling; DNA computing; parallel and distributed processing; scheduling; forecasting; circuit design; data mining; ant colony; coevolution; cultural algorithms ; evolutionary computation education; agent systems; breast cancer; neural nets; route and network planning; dynamic environments; fitness distributions; particle swarm; dynamic fitness; control systems; manufacturing optimization; quantum computing; IE/OR problems; and hybrid algorithms.
Subfile: C

15/7/50 (Item 9 from file: 2)
DIALOG(R)File 2:INSPEC

5998113 INSPEC Abstract Number: C9809-6110-032
Title: "Chromosome-Protein": a representation scheme
Author(s): Kai Zhao; Jue Wang
Author Affiliation: Inst. of Autom., Acad. Sinica, Beijing, China
Conference Title: Genetic Programming 1997 Proceedings of the Second Annual Conference p.343
Editor(s): Koza, J.R.; Deb, K.; Dorigo, M.; Fogel, D.B.; Garzon, M.; Iba, H.; Riolo, R.L.
Publisher: Morgan Kaufamann Publishers, San Francisco, CA, USA
Publication Date: 1997 Country of Publication: USA xviii+542 pp.
ISBN: 1 55860 483 9 Material Identity Number: XX97-01855
Conference Title: Proceedings of Genetic Programming 1997 Conference
Conference Date: 13-16 July 1997 Conference Location: Stanford, CA, USA
Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)
Abstract: When genetic programming (GP) is used in different design problems, different domain functions are provided. Trying to save these efforts and make a more general scheme, a library is used. A group of functions operating on the library and its elements as domain independent as possible are also suggested. This method is in accordance with a biological process called gene expression. Emulating gene expression, the Chromosome- Protein scheme is established. Two design problems, timetable and electrical circuit design, are taken as examples to illustrate the scheme. (0 Refs)
Subfile: C

15/7/51 (Item 10 from file: 2)
DIALOG(R)File 2:INSPEC

5897584 INSPEC Abstract Number: C9806-7330-034
Title: Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches
Author(s): Sung Jin Cho; Weifan Zheng; Tropsha, A.
Author Affiliation: Sch. of Pharm., North Carolina Univ., Chapel Hill, NC, USA
Journal: Journal of Chemical Information and Computer Sciences vol.38, no.2 p.259-68
Publisher: ACS,
Publication Date: March-April 1998 Country of Publication: USA
CODEN: JCISD8 ISSN: 0095-2338
SICI: 0095-2338(199803/04)38:2L.259:RCLD;1-4
Material Identity Number: J263-98002
U.S. Copyright Clearance Center Code: 0095-2338/98/$15.00
Document Number: S0095-2338(97)00094-2
Language: English Document Type: Journal Paper (JP)
Treatment: Theoretical (T)
Abstract: For pt. 1 see ibid., pp. 251-8. The authors have developed a novel strategy for rational design of targeted peptide libraries. The goal of this method is to select a subset of natural amino acids that are most likely to be present in active peptides for the synthesis of library. Two different protocols are employed where chemical structures of peptides are described either by topological indices or by a combination of physicochemical descriptors for individual amino acids. The selection of a peptide as a candidate for the targeted library is based either on its chemical similarity to a biologically active probe or on its biological activity predicted from a preconstructed quantitative structure-activity (QSAR) equation. The optimization of the library is achieved by means of genetic algorithms (GA). This method was tested by rational design of the library with bradykinin-potentiating activity. Twenty-eight bradykinin-potentiating pentapeptides were used as a training set for the development of a QSAR equation, and, alternatively, two active pentapeptides, VEWAK and VKWAP, were used as probe molecules. In each case, the frequency distribution of amino acids in the top 100 peptides suggested by the method resembles the frequency distribution of amino acids found in the active peptides. The results obtained after GA optimization also compared favorably with those obtained by the exhaustive analysis of all possible 3.2 million pentapeptides. (24 Refs)

Subfile: C
Copyright 1998, IEE

15/7/52 (Item 11 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2001 Institution of Electrical Engineers. All rts. reserv.

5809398 INSPEC Abstract Number: C9802-1240-004
Title: Nearly tight bounds on the learnability of evolution
Author(s): Ambainis, A.; Desper, R.; Farach, M.; Kannan, S.
Author Affiliation: Latvian State Univ., Riga, Latvia
Conference Title: Proceedings. 38th Annual Symposium on Foundations of Computer Science (Cat. No.97CB36150) p.524-33
Publisher: IEEE Comput. Soc, Los Alamitos, CA, USA
Publication Date: 1997 Country of Publication: USA xiii+606 pp.
ISBN: 0 8186 8197 7 Material Identity Number: XX98-00003
U.S. Copyright Clearance Center Code: 0272-5428/97/$10.00
Conference Title: Proceedings 38th Annual Symposium on Foundations of Computer Science
Conference Sponsor: IEEE Comput. Soc.; IEEE Comput. Soc. Tech. Committee on Math. Found. Comput
Conference Date: 20-22 Oct. 1997 Conference Location: Miami Beach, FL, USA
Language: English Document Type: Conference Paper (PA)
Treatment: Theoretical (T)
Abstract: Evolution is often modeled as a stochastic process which modifies DNA . One of the most popular and successful such processes are the Cavender-Farris (CF) trees, which are represented as edge weighted trees. The Phylogeny Construction Problem is that of, given kappa samples drawn from a CF tree, output a CF tree which is close to the original. Each CF tree naturally defines a random variable, and the gold standard for reconstructing such trees is the maximum likelihood estimator of this variable. This approach is notoriously computationally expensive. We show that a very simple algorithm, which is a variant on one of the most popular algorithms used by practitioners, converges on the true tree at a rate which differs from the optimum by a constant. We do this by analyzing upper and lower bounds for the convergence rate of learning very simple CF trees, and then show that the learnability of each CF tree is sandwiched between two such simpler trees. Our results rely on the fact that, if the right metric is used, the likelihood space of CF trees is smooth. (5 Refs)
Subfile: C
Copyright 1998, IEE

15/7/53 (Item 12 from file: 2)
DIALOG(R)File 2:INSPEC
(c) 2001 Institution of Electrical Engineers. All rts. reserv.

5558302 INSPEC Abstract Number: C9705-7330-309
Title: An object-oriented environment for artificial evolution of protein sequences: the example of rational design of transmembrane sequences
Author(s): Milik, M.; Skolnick, J.
Author Affiliation: Scripps Res. Instn., La Jolla, CA, USA
Conference Title: Evolutionary Programming IV. Proceedings of the Fourth Annual Conference on Evolutionary Programming p.603-13
Editor(s): McDonnell, J.R.; Reynolds, R.G.; Fogel, D.B.

Publisher: MIT Press, Cambridge, MA, USA
Publication Date: 1995 Country of Publication: USA xx+805 pp.
ISBN: 0 262 13317 2 Material Identity Number: XX96-02160
Conference Title: Proceedings of Fourth Annual Conference on Evolutionary Programming
Conference Date: 1-3 March 1995 Conference Location: San Diego, CA, USA
Language: English Document Type: Conference Paper (PA)
Treatment: Applications (A); Practical (P)
Abstract: A system is presented for generating peptide sequences with desirable properties using a combination of neural network and artificial evolution. The process is illustrated by an example of a practical problem of generating artificial transbilayer peptides . The peptides generated in the process of artificial evolution have the physico-chemical properties of transmembrane peptides, and form stable transmembrane structures in testing Monte Carlo simulations. The artificial evolution system is designed to emulate natural evolution; therefore it is of both practical and theoretical interest, both in terms of rational design of protein sequences and modeling of natural evolution of proteins. (3 Refs)
Subfile: C

15/7/54 (Item 13 from file: 2)
DIALOG(R)File 2:INSPEC

5323484 INSPEC Abstract Number: C9608-4240C-056
Title: DNA models and algorithms for NP-complete problems
Author(s): Bach, E.; Glaser, E.; Condon, A.; Tanguay, C.
Author Affiliation: Dept. of Comput. Sci., Wisconsin Univ., Madison, WI, USA
Conference Title: Proceedings Eleventh Annual IEEE Conference on Computational Complexity (Formerly: Structure in Complexity Theory Conference) (Cat. No.96CB35951) p.290-300
Editor(s): Homer, S.; Cai, J.-Y.
Publisher: IEEE Comput. Soc. Press, Los Alamitos, CA, USA
Publication Date: 1996 Country of Publication: USA x+307 pp.
ISBN: 0 8186 7386 9 Material Identity Number: XX96-01091
U.S. Copyright Clearance Center Code: 0 8186 7386 9/96/$05.00
Conference Title: Proceedings of Computational Complexity (Formerly Structure in Complexity Theory)
Conference Sponsor: IEEE Comput. Soc.; IEEE Comput. Soc. Tech. Committee on Math. Found. Comput.; ACM SIGACY; EATCS
Conference Date: 24-27 May 1996 Conference Location: Philadelphia, PA, USA
Language: English Document Type: Conference Paper (PA)
Treatment: Theoretical (T)
Abstract: A goal of research on DNA computing is to solve problems that are beyond the capabilities of the fastest silicon-based supercomputers. Adleman and Lipton present exhaustive search algorithms for 3Sat and 3-Coloring, which can only be run on small instances and hence are not practical. In this paper, we show how improved algorithms can be developed for the 3-Coloring and Independent Set problems. Our algorithms use only the DNA operations proposed by Adleman and Lipton, but combine them in more powerful ways, and use polynomial preprocessing on a standard computer to

tailor them to the specific instance to be solved. The main contribution of this paper is a more general model of DNA algorithms than that proposed by Lipton. We show that DNA computation for NP-complete problems can do more than just exhaustive search. Further research in this direction will help determine whether or not DNA computing is viable for NP-hard problems. A second contribution is the first analysis of errors that arise in generating the solution space for DNA computation. (10 Refs)
Subfile: C

15/7/55 (Item 14 from file: 2)
DIALOG(R)File 2:INSPEC

4951398 INSPEC Abstract Number: A9511-8710-006, C9506-1290L-064
Title: Studying genotype-phenotype interactions: a model of the evolution of the cell regulation network
Author(s): Chiva, E.; Tarroux, P.
Author Affiliation: Lab. de Biol. et Physiol. de Developpement, Ecole Normale Superieure, Paris, France
p.26-35
Editor(s): Davidor, Y.; Schwefel, H.-P.; Manner, R.
Publisher: Springer-Verlag, Berlin, Germany
Publication Date: 1994 Country of Publication: West Germany xv+642 pp.
ISBN: 3 540 58484 6
Conference Title: Parallel Problem Solving from Nature - PPSN III International Conference on Evolutionary Computation The Third Conference on Parallel Problem Solving from Nature
Conference Date: 9-14 Oct. 1994 Conference Location: Jerusalem, Israel
Language: English Document Type: Conference Paper (PA)
Treatment: Theoretical (T)
Abstract: A new model of the interactions between genotype and phenotype is presented. These interactions are underlied by the activity of the cell regulation network. This network is modeled by a continuous recurrent automata network, which describes both direct and indirect interactions between proteins. To mimic evolutionary processes, a particular genetic algorithm is used to simulate the environmental influences on the interactions between proteins . The fitness function is designed to select systems that are robust to transient environmental perturbations, thus exhibiting homeostasy, and that respond in an adapted way to lasting perturbations by a radical change in their behavior. We show that by evaluating the phenotypic response of the system, one can select networks that exhibit interesting dynamical properties, which allows to consider a biological system from a global perspective, taking into account its structure, its behavior and its ontogenetic development. This model provides a new biological metaphor in which the cell is considered as a cybernetic system that can be programmed using a genetic algorithm . (21 Refs)
Subfile: A C

15/7/56 (Item 15 from file: 2)
DIALOG(R)File 2:INSPEC

4695943 INSPEC Abstract Number: C9408-6115-005
Title: Development needs for diverse genetic algorithm design
Author(s): Kingdon, J.; Dekker, L.
Author Affiliation: Dept. of Comput. Sci., Univ. Coll. London, UK
Conference Title: IEE Colloquium on 'Applications of Genetic Algorithms' (Digest No.1994/067) p.3/1-11
Publisher: IEE, London, UK
Publication Date: 1994 Country of Publication: UK 38 pp.
Conference Title: IEE Colloquium on 'Applications of Genetic Algorithms' (Digest No.1994/067)
Conference Sponsor: IEE
Conference Date: 15 March 1994 Conference Location: London, UK
Language: English Document Type: Conference Paper (PA)
Treatment: Practical (P)
Abstract: This paper describes the development of an object-oriented parallel programming environment for genetic algorithms . This work, carried out as part of the ESPRIT III initiative PAPAGENA, intends to promote, develop and demonstrate the effectiveness of genetic algorithm (GA) and parallel genetic algorithm (PGA) techniques in a variety of real-world application domains. Central to this task is the development of a general-purpose programming environment for both parallel and sequential genetic algorithms . GAME (Genetic Algorithm Manipulation Environment) will offer extensive tools for the design, configuration and monitoring of GA applications. This paper gives an overview of the design philosophy behind GAME, indicating the types of service and facilities the finished product will offer. Intrinsic to the design is the provision of an extensive multi-levelled GA-specific library, offering GA and PGA applications, algorithms and operators. This will allow application developers the facilities to rapidly customise, configure and test novel GA and PGA designs. To sketch the types of application to be housed in GAME, a description of the applications currently under development within this project is also included. These range from finance through economic modelling, to protein structure prediction. Key design requirements for GAME are versatility, together with flexibility. For this reason GAME has been designed to run within both Sun OS and PC DOS operating system, with or without parallel support. (26 Refs)
Subfile: C

15/7/57 (Item 16 from file: 2)
DIALOG(R)File 2:INSPEC

4665062 INSPEC Abstract Number: A9411-0130C-070, C9406-1290L-041
Title: IEE Colloquium on 'Molecular Bioinformatics' (Digest No.1994/029)
Publisher: IEE, London, UK
Publication Date: 1994 Country of Publication: UK 43 pp.
Conference Title: IEE Colloquium on 'Molecular Bioinformatics' (Digest No.1994/029)
Conference Sponsor: IEE
Conference Date: 28 Feb. 1994 Conference Location: London, UK
Language: English Document Type: Conference Proceedings (CP)
Abstract: The following topics were dealt with: machine learning application to protein structure and drug design ; protein design methods and algorithms; cellular information system models; self-organising feature maps in singular domains identification; ribonucleic acid secondary

structure determination; DNA database building; genetics sequencing and genetic algorithm .
Subfile: A C

15/7/58 (Item 17 from file: 2)
DIALOG(R)File 2:INSPEC

04197573 INSPEC Abstract Number: A9217-8715B-008
Title: Protein folding and deterministic chaos: limits of protein folding simulations and calculations
Author(s): Bohm, G.
Author Affiliation: Regensburg Univ., Germany
Journal: Chaos, Solitons and Fractals vol.1, no.4 p.375-82
Publication Date: 1991 Country of Publication: UK
CODEN: CSFOEH ISSN: 0960-0779
U.S. Copyright Clearance Center Code: 0960-0779/91/$3.00+00
Language: English Document Type: Journal Paper (JP)
Treatment: Theoretical (T)
Abstract: The aim of the present work is to suggest that protein folding is a highly complex process which generally cannot be simulated on digital computers. This limitation is not due to the nonavailability of computing resources, as it has been suggested previously, but to the lack of exact force field parameters; it is obviously impossible to quantify parameter(s) for any deterministic algorithm with sufficient accuracy to describe the dynamics of protein folding. Molecular dynamics simulations on crambin, a small protein with 46 amino acids whose three-dimensional structure is known, suggest the native state to be a 'fixed attractor'. The results show that any ab initio calculation of protein structure must fail if the folding process of a protein is controlled by a kinetic process, i.e. when the native state is the kinetically accessible minimum on the energy hyperspace but not the thermodynamically possible global minimum. In this case only non-dynamical methods like pattern recognition or database processing (knowledge-based approaches) can provide a reasonable three-dimensional structure. Novel computational methods like genetic algorithms and neural network methods may be more valuable for the design and description of protein structures than the traditional force-field based algorithmic methods used to date. Knowledge-based modelling is therefore the most promising method to date to deduce the structure of an unknown protein from the sequence information solely. (19 Refs)
Subfile: A

15/7/59 (Item 1 from file: 8)
DIALOG(R)File 8:Ei Compendex(R)

05745450 E.I. No: EIP00125457380
Title: DNA genetic algorithm for design of the generalized membership-type Takagi-Sugeno fuzzy control system
Author: Ding, Yongsheng; Ren, Lihong
Corporate Source: China Textile Univ, Shanghai, China
Conference Title: 2000 IEEE International Conference on Systems, Man and Cybernetics
Conference Location: Nashville, TN, USA Conference Date:

20001008-20001011
Sponsor: IEEE
E.I. Conference No.: 57756
Source: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics v 5 2000. IEEE, Piscataway, NJ, USA,00CB37166. p 3862-3867
Publication Year: 2000
CODEN: PICYE3 ISSN: 0884-3627
Language: English
Document Type: CA; (Conference Article) Treatment: T; (Theoretical)
Journal Announcement: 0102W2
Abstract: In this paper we propose a new DNA-based genetic algorithm (DNA -GA) to optimize the design parameters of a generalized membership-type Takagi-Sugeno fuzzy controller (GTSFC). The GTSFC employs TS fuzzy rules with linear consequent, e** minus vertical bar ax plus b vertical bar **c-type input fuzzy sets containing almost arbitrary continuous input fuzzy sets, Zadeh fuzzy logic AND operation, and the widely-used centroid defuzzier. The GTSFC is proved to be a nonlinear PI controller with variable gains. The optimized design parameters are the input fuzzy sets and the linear consequent of the rules. The DNA-GA uses DNA encoding method stemmed from the structure of the biological DNA to encode the design parameters of the GTSFC. The genetic operators of the method are based on the DNA genetic operations. The encoding method can significantly shorten the code length of DNA chromosomes and is suitable for complex knowledge representation. As a demonstration we show how to implement the new method to optimize the design parameters of the GTSFC to control a nonlinear system. Computer simulation results indicate that the performance of the designed fuzzy controller is satisfactory. (Author abstract) 23 Refs.

15/7/60 (Item 2 from file: 8)
DIALOG(R)File 8:Ei Compendex(R)

05638153 E.I. No: EIP00085296208
Title: Synthesis of purification tags for optimal downstream processing
Author: Steffens, M.A.; Fraga, E.S.; Bogle, I.D.L.
Corporate Source: Univ Coll London, London, Engl
Conference Title: 7th International Symposium on Process Systems Engineering
Conference Location: Keystone, CO, USA Conference Date: 19000716-19000721
Sponsor: CACHE Corporation; AICHE
E.I. Conference No.: 57163
Source: Computers and Chemical Engineering v 24 n 2 Jul 2000. p 717-720
Publication Year: 2000
CODEN: CCENDW ISSN: 0098-1354
Language: English
Document Type: JA; (Journal Article) Treatment: X; (Experimental)
Journal Announcement: 0010W1
Abstract: Downstream purification yields and costs can be significantly improved by genetically modifying a product protein to alter its physical properties. Purification tags, or sequences of amino acids which are genetically fused onto the product protein, have been extensively researched and used for this purpose. However, all of the work in this area has involved specific purification tags, which have advantages in certain situations but are not generically optimal. In this work we describe the

development and application of a combinatorial approach which can generate the best sequence of amino acids for a particular product protein. The aim of the work is to develop an algorithm which generates sequences of amino acids (i.e. purification tags) which may be fused onto a specific product protein and used to aid in subsequent purification steps. The work is integrated into a previously developed bioprocess synthesis algorithm, which enables us to compare the flowsheets generated with and without purification tags for a case study system. The results show that the approach allows us to synthesize flowsheets with fewer units, higher yields and lower costs. (Author abstract) 12 Refs.

15/7/61 (Item 3 from file: 8)
DIALOG(R)File 8:Ei Compendex(R)

05618044 E.I. No: EIP00085269381
Title: Investigations into integrated location management in mobile multimedia networks
Author: Ponnekanti, Seshaiah; Prasad, Ramjee
Corporate Source: Univ of Hertfordshire, Herts, UK
Conference Title: 1st International Conference on '3G Mobile Communication Technologies'
Conference Location: London, UK Conference Date: 19000327-19000329
E.I. Conference No.: 57135
Source: IEE Conference Publication n 471 2000. IEE, Stevenage, Engl. p 294-300
Publication Year: 2000
CODEN: IECPB4 ISSN: 0537-9989
Language: English
Document Type: CA; (Conference Article) Treatment: T; (Theoretical)
Journal Announcement: 0009W2
Abstract: Advanced software protocols for multimedia packet networks are currently designed tested and, in some cases, fielded by a variety of organizations. Additionally, variety of solutions are required to provide next -generation , highly-programmable technology solutions to eventually provide reliable, efficient, and secure communications of multimedia traffic over rapidly-deployable mobile adhoc networks to pave the way for seamless extension of the internet. In this context, it is recognized that user location identification is considered by the service providers as one of the most value-added services for the end-users. The focus of the research in these areas is to allow protocol software and the location technology solutions to be integrated in a flexible manner into the mobile multimedia networks through API (Application Programming Interface) framework. Various location strategies needed to be analyzed to derive candidate solutions for integrated location management. In this paper, the background preparation to derive a separate smart antenna API is defined to eventually incorporate it into the distributed network architecture. Preliminary results from the real-data smart antenna experiments are examined to find out the efficacy of the technique. (Author abstract) 2 Refs.

15/7/62 (Item 4 from file: 8)
DIALOG(R)File 8:Ei Compendex(R)

05424896 E.I. No: EIP99114912379
Title: High performance phylogenetic inference
Author: Clement, Mark; Snell, Quinn; Judd, Glenn; Whiting, Michael
Corporate Source: Brigham Young Univ, Provo, UT, USA
Conference Title: Proceedings of the 1999 8th IEEE International Symposium on High Performance Distributed Computing - HPDC-8
Conference Location: Redondo Beach, CA, USA Conference Date: 19990803-19990806
Sponsor: IEEE Computer Society; USC-Information Science Institute; UCSD; NPACI; et al.
E.I. Conference No.: 55898
Source: IEEE International Symposium on High Performance Distributed Computing, Proceedings 1999. p 335-336
Publication Year: 1999
CODEN: PIDCFB ISSN: 1082-8907
Language: English
Document Type: JA; (Journal Article) Treatment: G; (General Review)
Journal Announcement: 0001W2
Abstract: Phylogenetic analysis is an integral part of many biological research programs. Researchers are now commonly generating many DNA sequences from many individuals, thus creating very large data sets. The ability to the data has not kept pace with data generation, and phylogenetics has reached a crossroads where the data cannot be analyzed effectively. The chief challenge of phylogenetic systematics in the next century will be to develop algorithms and search strategies to be effectively analyze data sets. The crux of the computational problem is that the actual landscape of possible topologies can be extraordinarily difficult to evaluate with large data sets. 6 Refs.

15/7/63 (Item 5 from file: 8)
DIALOG(R)File 8:Ei Compendex(R)

04809907 E.I. No: EIP97093806959
Title: Proceedings of the 1997 IEEE International Conference on Neural Networks. Part 1 (of 4)
Author: Anon (Ed.)
Conference Title: Proceedings of the 1997 IEEE International Conference on Neural Networks. Part 1 (of 4)
Conference Location: Houston, TX, USA Conference Date: 19970609-19970612
Sponsor: IEEE
E.I. Conference No.: 46924
Source: IEEE International Conference on Neural Networks - Conference Proceedings v 1 1997. IEEE, Piscataway, NJ, USA,97CB36109. 616p
Publication Year: 1997
CODEN: ICNNF9
Language: English
Document Type: CP; (Conference Proceedings) Treatment: A; (Applications); G; (General Review); T; (Theoretical)
Journal Announcement: 9710W4
Abstract: The proceedings contains 116 papers on neural networks from the 1997 IEEE International Conference. Topics discussed include: adaptive critic design; time delay spectrometry; option pricing; ultrashort laser pulse characterization; partial discharge analysis; real time plasma boundary reconstruction; flow field computation; anterior wall myocardial

infarction; predictive medicine; prosthetics design ; large-scale protein family identification; bird identification; polymer product online identification; wind power generation; wireless telephony; reactive ion etching; fatigue damage prediction; automatic structure search; thalamic electric oscillator; and intelligent sales forecasting system.

15/7/64 (Item 6 from file: 8)
DIALOG(R)File 8:Ei Compendex(R)

04697556 E.I. No: EIP97053658270
Title: On the power of circular splicing systems and DNA computability
Author: Yokomori, Takashi; Kobayashi, Satoshi; Ferretti, Claudio
Corporate Source: Univ of Electro-Communications, Tokyo, Jpn
Conference Title: Proceedings of the 1997 IEEE International Conference on Evolutionary Computation, ICEC'97
Conference Location: Indianapolis, IN, USA Conference Date: 19970413-19970416
Sponsor: IEEE
E.I. Conference No.: 46380
Source: Proceedings of the IEEE Conference on Evolutionary Computation, ICEC 1997. IEEE, Piscataway, NJ, USA,97TH8283. p 219-224
Publication Year: 1997
CODEN: 001660
Language: English
Document Type: CA; (Conference Article) Treatment: T; (Theoretical)
Journal Announcement: 9707W2
Abstract: From a biological motivation of interactions between linear and circular DNA sequences, we propose a new type of splicing models called circular H systems and show that they have the same computational power as Turing machines. It is also shown that there effectively exists a universal circular H system which can simulate any circular H system with the same terminal alphabet, which strongly suggests a feasible design for a DNA computer based on circular splicing. (Author abstract) 21 Refs.

15/7/65 (Item 7 from file: 8)
DIALOG(R)File 8:Ei Compendex(R)

04223493 E.I. No: EIP94112425013
Title: Hybrid genetic algorithm application to a genetics sequencing problem
Author: Walker, J.D.; File, P.E.; Miller, C.J.; Samson, W.B.
Corporate Source: Dundee Inst of Technology, Scotl
Conference Title: Colloquium on Molecular Bioinformatics
Conference Location: London, UK Conference Date: 19940228
Sponsor: Professional Group C4
E.I. Conference No.: 21231
Source: IEE Colloquium (Digest) n 029 Feb 28 1994. IEE, Stevenage, Engl. p 7/1-7/12
Publication Year: 1994
CODEN: DCILDN ISSN: 0963-3308
Language: English
Document Type: CA; (Conference Article) Treatment: G; (General Review)
Journal Announcement: 9510W1

Abstract: Currently, geneticists are analysing the structure of the DNA of various organisms to determine the location and sequence of genes. A technique which has played a major role in this process is the use of restriction enzymes and radioactively labelled probes to generate maps of the DNA . Building restriction maps from the results of probed partial digest experiments is a time-consuming and lengthy activity which relies on human judgement of inexact data. Map building is an example of a difficult sequencing problem which requires some form of search to find a good solution from a large problem space of feasible solutions. This paper describes the development of a hybrid genetic algorithm (HGA) suitable for tackling the problem. The results of applying the HGA to a set of map data are presented. (Author abstract) 19 Refs.

15/7/66 (Item 1 from file: 65)
DIALOG(R)File 65:Inside Conferences

03332711 INSIDE CONFERENCE ITEM ID: CN035222194
DNA Genetic Algorithms for Design of Fuzzy Systems
Ren, L.; Ding, Y.; Shao, S.
CONFERENCE: Fuzzy systems-International conference; 9th
IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS, 2000; VOL 2 P: 1005-1008
IEEE, 2000
ISSN: 1098-7584 ISBN: 0780358783; 0780358775; 0780363248; 0780363248
LANGUAGE: English DOCUMENT TYPE: Conference Preprinted papers
CONFERENCE SPONSOR: IEEE
CONFERENCE LOCATION: San Antonio, TX
CONFERENCE DATE: May 2000
NOTE:
Also known as FUZZ-IEEE 2000. IEEE cat no 00CH37063

15/7/67 (Item 2 from file: 65)
DIALOG(R)File 65:Inside Conferences

02671158 INSIDE CONFERENCE ITEM ID: CN027807918
Designing Genetic Algorithm Fitness Functions for Protein Tertiary Structure Prediction
Yap, A.; Cosic, I.
CONFERENCE: Protein structure and function-Annual Lorne conference; 22nd
ANNUAL LORNE CONFERENCE ON PROTEIN STRUCTURE AND FUNCTION, 1997; 22nd
P: C58
(np), 1997
ISSN: 1034-3180
LANGUAGE: English DOCUMENT TYPE: Conference Abstracts and programme
CONFERENCE LOCATION: Lorne, Australia
CONFERENCE DATE: Feb 1997 (199702) (199702)
NOTE:
Also known as the 22nd Lorne protein conference, 1997

15/7/68 (Item 1 from file: 144)
DIALOG(R)File 144:Pascal

14661617 PASCAL No.: 00-0334345
A surface-based DNA algorithm for the expansion of symbolic determinants
Parallel and distributed processing : Cancun, 1-5 May 2000
QIU Z F; MI LU
ROLIM Jose, ed
Department of Electrical Engineering, Texas A&M University, College Station, Texas 77843-3128, United States
15 IPDPS 2000 workshops (Cancun MEX) 2000-05-01
Journal: Lecture notes in computer science, 2000, 1800 653-659
ISBN: 3-540-67442-X ISSN: 0302-9743 Availability: INIST-16343; 354000080064550810
No. of Refs.: 22 ref.
Document Type: P (Serial); C (Conference Proceedings) ; A (Analytic)
Country of Publication: Germany
Language: English

In the past few years since Adleman's pioneering work on solving the HPP(Hamiltonian Path Problem) with a DNA-based computer (1), many algorithms have been designed on solving NP problems. Most of them are in the solution bases and need some error correction or tolerance technique in order to get good and correct results (3) (7) (9) (11) (21) (22). The advantage of surface-based DNA computing technique, with very low error rate, has been shown many times (12) (18) (17) (20) over the solution based DNA computing, but this technique has not been widely used in the DNA computer algorithms design . This is mainly due to the restriction of the surface-based technique comparing with those methods using the DNA strands in solutions. In this paper, we introduce a surface-based DNA computing algorithm for solving a hard computation problem: expansion of symbolic determinants given their patterns of zero entries. This problem is well-known for its exponential difficulty. It is even more difficult than evaluating determinants whose entries are merely numerical (15). We will show how this problem can be solved with the low error rate surface-based DNA computer using our naive algorithm.

15/7/69 (Item 2 from file: 144)
DIALOG(R)File 144:Pascal

13842823 PASCAL No.: 99-0019086
A fast algorithm for discovering optimal string patterns in large text databases
ALT '98 : algorithmic learning theory : Otzenhausen, 8-10 October 1998
ARIMURA H; WATAKI A; FUJINO R; ARIKAWA S
RICHTER Michael M, ed; SMITH Carl H, ed; WIEHAGEN Rolf, ed; ZEUGMANN Thomas, ed
Department of Informatics, Kyushu University, Hakozaki 6-10-1, Fukuoka, 812-8581, Japan
Algorithmic learning theory. International workshop, 9 (Otzenhausen DEU) 1998-10-08
Journal: Lecture notes in computer science, 1998, 1501 247-261
ISBN: 3-540-63577-7 ISSN: 0302-9743 Availability: INIST-16343; 354000070132130190
No. of Refs.: 28 ref.
Document Type: P (Serial); C (Conference Proceedings) ; A (Analytic)
Country of Publication: Germany; United States

Language: English

We consider a data mining problem in a large collection of unstructured texts based on association rules over subwords of texts. A two-words association pattern is an expression such as (TATA, 30, AGGAGGT) Rightarrow C that expresses a rule that if a text contains a subword TATA followed by another subword AGGAGGT with distance no more than 30 letters then a property C will hold with high probability. The optimized confidence pattern problem is to compute frequent patterns (alpha , kappa , beta) that optimize the confidence with respect to a given collection of texts. Although this problem is solved in polynomial time by a straightforward algorithm that enumerates all the possible patterns in time O(n SUP 5), we focus on the development of more efficient algorithms that can be applied to large text databases. We present an algorithm that solves the optimized confidence pattern problem in time O(max(kappa ,m)n SUP 2) and space O(kappa n) where m and n are the number and the total length of classification examples, respectively, and kappa is a small constant around 30 similar 50. This algorithm combines the suffix tree data structure in combinatorial string matching and the orthogonal range query technique in computational geometry for fast computation. Furthermore for most random texts like DNA sequences, we show that a modification of the algorithm runs very efficiently in time O(kappa n log SUP 3 n) and space O(kappa n). We also discuss some heuristics such as sampling and pruning as practical improvement. Then, we evaluate the efficiency and the performance of the algorithm with experiments on genetic sequences. A relationship with efficient Agnostic PAC-learning is also discussed.

15/7/70 (Item 3 from file: 144)
DIALOG(R)File 144:Pascal

12389390 PASCAL No.: 96-0036514

Genome sequence comparison and scenarios for gene rearrangements : a test cause

SRIDHAR HANNENHALLI; CHAPPEY C; KOONIN E V; PEVZNER P A

Pennsylvania State univ., dep. computer sci. eng., University Park PA 16802, USA

Journal: Genomics ; (San Diego, CA), 1995, 30 (2) 299-311

ISSN: 0888-7543 Availability: INIST-21389; 354000059060790210

No. of Refs.: 1 p.1/2

Document Type: P (Serial) ; A (Analytic)

Country of Publication: USA

Language: English

As large portions of related genomes are being sequenced, methods for comparing complete or nearly complete genomes, as opposed to comparing individual genes, are becoming progressively more important. A major, widespread phenomenon in genome evolution is the rearrangement of genes and gene blocks. There is, however, no consistent method for genome sequence comparison combined with the reconstruction of the evolutionary history of highly rearranged genomes. We developed a schema for genome sequence comparison that includes three successive steps : (i) comparison of all proteins encoded in different genomes and generation of genomic similarity plots ; (ii) construction of an alphabet of conserved genes and gene blocks ; and (iii) generation of most parsimonious genome rearrangement scenarios. The approach is illustrated by a comparison of the herpesvirus genomes that constitute the largest set of relatively long,

complete genome sequences available to date. Herpesviruses have from 70 to about 200 genes ; comparison of the amino acid sequences encoded in these genes results in an alphabet of about 30 conserved genes comprising 7 conserved blocks that are rearranged in the genomes of different herpesviruses. Algorithms to analyze rearrangements of multiple genomes were developed and applied to the derivation of most parsimonious scenarios of herpesvirus evolution under different evolutionary models. The developed approaches to genome comparison will be applicable to the comparative analysis of bacterial and eukaryotic genomes as soon as their sequences become available.

15/7/71 (Item 4 from file: 144)
DIALOG(R)File 144:Pascal

10467764 PASCAL No.: 92-0671253
Atomic resolution structures of DNA and DNA modified by carcinogens
HINGERTY B E; BROYDE S
Oak Ridge National Laboratory, Oak Ridge TN 37831, USA
Journal: (The) International journal of supercomputer applications; (The) International journal of supercomputer applications, 1990, 4 (3) 11-21
ISSN: 0890-2720 Availability: INIST-21391; 354000004338710030
No. of Refs.: 17 ref.
Document Type: P (Serial) ; A (Analytic)
Country of Publication: USA
Language: English

15/7/72 (Item 5 from file: 144)
DIALOG(R)File 144:Pascal

09976959 PASCAL No.: 92-0193903
A combinatorial description of the closest tree algorithm for finding evolutionary trees
HENDY M D
Massey univ., dep. mathematics statistics, Palmerston North, New Zealand
Journal: Discrete mathematics, 1991, 96 (1) 51-58
ISSN: 0012-365X CODEN: DSMHA4 Availability: INIST-15322; 354000022509350030
No. of Refs.: 8 ref.
Document Type: P (Serial) ; A (Analytic)
Country of Publication: Netherlands
Language: English
The closest tree algorithm for estimating the evolutionary history of n species, from a set of homologous DNA or RNA sequences is designed to avoid the problem of inconsistency inherent in current methods. The algorithm, as previously described, required O(n SUP n 2 SUP n) steps, making it impractical for values of n>10. In this paper, a new description of the algorithm is given, exploiting a combinatorial inverse pair relationship. As a consequence, the algorithm can be improved in efficiency, to be O(n2 SUP n) for some classes of sequences. This improvement makes the algorithm practical for problems of involving up to n=20 species

15/7/73 (Item 1 from file: 35)
DIALOG(R)File 35:Dissertation Abstracts Online

01605180 ORDER NO: AAD98-07318
DETERMINATION OF 3D STRUCTURE OF SURAMIN IN SOLUTION AND MODEL OF INTERACTION OF SURAMIN WITH BASIC FIBROBLAST AND PLATELET-DERIVED GROWTH FACTORS (ANTICANCER AGENTS)

Author: MARCUS, EMIL
Degree: PH.D.
Year: 1997
Corporate Source/Institution: STATE UNIVERSITY OF NEW YORK AT BUFFALO (0656)
Adviser: ROBER REIN
Source: VOLUME 58/08-B OF DISSERTATION ABSTRACTS INTERNATIONAL. PAGE 4098. 248 PAGES

This research work focuses on the determination for the first time of the 3D solution structure of potent anticancer drug suramin, investigation of its conformational properties, and use of this structure to propose molecular models for its interaction with and inhibitory effect exerted upon two growth factors: bFGF and PDGF-BB.

Joint use of 2D NMR and modeling techniques revealed C$\sb2$ symmetry and an unusually high segmental flexibility at the level of the second pair of secondary amides that anchor the two naphthalene systems. Free rotations about the C(phenyl)-C(carbonyl) and C(naphthalene)-C(amide) bonds and rotation of each sulfonate group about C(naphthalene)-S(sulfone) bonds appear to direct the orientation of these charged groups, thus accounting for the ability to interact non-specifically with a variety of structurally and functionally diverse proteins.

Docking simulations in the Amber force field with a genetic algorithm (GA) were carried out by allowing the suramin molecule to adjust its conformation to the protein surface (bFGF and PDGF), while, at the same time, permitting the exposed basic residues to adjust their molecular conformation to facilitate the formation of salt bridges and H-bonds with suramin. In both cases, two suramin molecules "engulf" one molecule of protein . GA generated docking scores seem to indicate good shape complementary associated with sufficient polar contacts to stabilize the interactions between suramin and each of the two proteins. Only two sulfone functions from each naphthalene group participate in the ligation process. The two complexes differ in that, in view of its internal molecular flexibility suramin is "capable" and actually "uses" two different conformers to dock to the growth factors: "open stance horse-shoe" for bFGF, and "closed stance horse-shoe" for PDGF.

Both models are consistent with most experimentally determined properties (such as the 2:1 binding stoichiometry), and, account for the molecular mechanisms responsible for the inhibitory effect suramin exerts upon these factors, essentially by preventing attachment of cognate receptors.

Results of these structure-biological investigations may be used to design suramin analogs with increased binding specificity to selected cytokines such as bFGF and PDGF, thus improving the therapeutic vs toxicity ratio, and, consequently, the efficiency of this pharmaceutical.

15/7/74 (Item 1 from file: 5)

DIALOG(R)File 5:Biosis Previews(R)

11709467 BIOSIS NO.: 199800491198
Peptide design by artificial neural networks and computer-based evolutionary search.
AUTHOR: Schneider Gisbert(a); Schroedl Wieland; Wallukat Gerd; Mueller Johannes; Nissen Eberhard; Roenspeck Wolfgang; Wrede Paul; Kunze Rudolf
AUTHOR ADDRESS: (a)F. Hoffmann-La Roche Ltd., Pharm. Div., Mol. Design Bioinf., CH-4070 Basel**Switzerland
JOURNAL: Proceedings of the National Academy of Sciences of the United States of America 95 (21):p12179-12184 Oct. 13, 1998
ISSN: 0027-8424
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: A technique for systematic peptide variation by a combination of rational and evolutionary approaches is presented. The design scheme consists of rive consecutive steps: (i) identification of a "seed peptide " with a desired activity, (ii) generation of variants selected from a physicochemical space around the seed peptide, (iii) synthesis and testing of this biased library, (iv) modeling of a quantitative sequence-activity relationship by an artificial neural network, and (v) de novo design by a computer-based evolutionary search in sequence space using the trained neural network as the fitness function. This strategy was successfully applied to the identification of novel peptides that fully prevent the positive chronotropic effect of anti-beta1-adrenoreceptor autoantibodies from the serum of patients with dilated cardiomyopathy. The seed peptide, comprising 10 residues, was derived by epitope mapping from an extracellular loop of human beta1-adrenoreceptor. A set of 90 peptides was synthesized and tested to provide training data for neural network development. De novo design revealed peptides with desired activities that do not match the seed peptide sequence. These results demonstrate that computer-based evolutionary searches can generate novel peptides with substantial biological activity.

15/7/75 (Item 2 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)

09938510 BIOSIS NO.: 199598393428
Automatic protein de novo design by genetic algorithm and 3D profile.
AUTHOR: Lai Luhua(a); Guo Feng; Xu Xiaojie; Tang Youqi
AUTHOR ADDRESS: (a)Inst. Physical Chem., Peking Univ., Beijing 100871** China
JOURNAL: Journal of Biomolecular Structure & Dynamics 12 (6):pA128 1995
CONFERENCE/MEETING: Ninth Conversation in Biomolecular Stereodynamics June 20-24, 1995
ISSN: 0739-1102
RECORD TYPE: Citation
LANGUAGE: English

15/7/76 (Item 1 from file: 73)
DIALOG(R)File 73:EMBASE

07859945 EMBASE No: 1999333727
De novo design and structural characterization of proteins and metalloproteins
DeGrado W.F.; Summa C.M.; Pavone V.; Nastri F.; Lombardi A.
W.F. DeGrado, Johnson Research Foundation, Dept. of Biochemistry and Biophysics, Univ. of Pennsylvania Medical School, Philadelphia, PA 19104-6059 United States
AUTHOR EMAIL: wdegrado@mail.med.upenn.edu
Annual Review of Biochemistry (ANNU. REV. BIOCHEM.) (United States) 1999, 68/- (779-819)
CODEN: ARBOA ISSN: 0066-4154
DOCUMENT TYPE: Journal; Review
LANGUAGE: ENGLISH SUMMARY LANGUAGE: ENGLISH
NUMBER OF REFERENCES: 309

De novo protein design has recently emerged as an attractive approach for studying the structure and function of proteins. This approach critically tests our understanding of the principles of protein folding; only in de novo design must one truly confront the issue of how to specify a protein's fold and function. If we truly understand proteins , it should be possible to design receptors, enzymes, and ion channels from scratch. Further, as this understanding evolves and is further refined, it should be possible to design proteins and biomimetic polymers with properties unprecedented in nature.

15/7/77 (Item 2 from file: 73)
DIALOG(R)File 73:EMBASE

06836935 EMBASE No: 1997119445
Development and validation of a genetic algorithm for flexible docking
Jones G.; Willett P.; Glen R.C.; Leach A.R.; Taylor R.
G. Jones, Department of Information Studies, University of Sheffield, Sheffield S10 2TN United Kingdom
Journal of Molecular Biology (J. MOL. BIOL.) (United Kingdom) 1997, 267/3 (727-748)
CODEN: JMOBA ISSN: 0022-2836
DOCUMENT TYPE: Journal; Article
LANGUAGE: ENGLISH SUMMARY LANGUAGE: ENGLISH
NUMBER OF REFERENCES: 87

Prediction of small molecule binding modes to macromolecules of known three-dimensional structure is a problem of paramount importance in rational drug design (the 'docking' problem). We report the development and validation of the program GOLD (Genetic Optimisation for Ligand Docking). GOLD is an automated ligand docking program that uses a genetic algorithm to explore the full range of ligand conformational flexibility with partial flexibility of the protein, and satisfies the fundamental requirement that the Ligand must displace loosely bound water on binding. Numerous enhancements and modifications have been applied to the original technique resulting in a substantial increase in the reliability and the

applicability of the algorithm. The advanced algorithm has been tested on a dataset of 100 complexes extracted from the Brookhaven Protein DataBank. When used to dock the ligand back into the binding site, GOLD achieved a 71% success rate in identifying the experimental binding mode.

15/7/78 (Item 3 from file: 73)
DIALOG(R)File 73:EMBASE

06795021 EMBASE No: 1997076524
Finding protein-binding sites in DNA sequences: The next generation
Frech K.; Quandt K.; Werner T.
K. Frech, GSF-Nat Res Center Environment/Hlth, Institute of Mammalian Genetics, Ingolstadter Landstrasse 1, D-85764 Neuherbergg Germany
Trends in Biochemical Sciences (TRENDS BIOCHEM. SCI.) (United Kingdom) 1997, 22/3 (103-104)
CODEN: TBSCD ISSN: 0968-0004
PUBLISHER ITEM IDENTIFIER: S0968000497010062
DOCUMENT TYPE: Journal; Short Survey
LANGUAGE: ENGLISH
NUMBER OF REFERENCES: 9

15/7/79 (Item 4 from file: 73)
DIALOG(R)File 73:EMBASE

06462815 EMBASE No: 1996128851
Towards meeting the Paracelsus Challenge: The design, synthesis, and characterization of paracelsin-43, an alpha-helical protein with over 50% sequence identity to an all-beta protein
Jones D.T.; Moody C.M.; Uppenbrink J.; Viles J.H.; Doyle P.M.; Harris C.J.; Pearl L.H.; Sadler P.J.; Thornton J.M.
Biomolec. Structure/Modelling Unit, Biochemistry/Molecular Biology Dept., University College, Gower Street,London WC1E 6BT United Kingdom
Proteins: Structure, Function and Genetics (PROTEINS STRUCT. FUNCT. GENET.) (United States) 1996, 24/4 (502-513)
CODEN: PSFGE ISSN: 0887-3585
DOCUMENT TYPE: Journal; Article
LANGUAGE: ENGLISH SUMMARY LANGUAGE: ENGLISH

In response to the Paracelsus Challenge (Rose and Creamer, Proteins, 19:1- 3, 1994), we present here the design, synthesis, and characterization of a helical protein, whose sequence is 50% identical to that of an all-beta protein. The new sequence was derived by applying an inverse protein folding approach, in which the sequence was optimized to 'fit' the new helical structure, but constrained to retain 50% of the original amino acid residues. The program utilizes a genetic algorithm to optimize the sequence, together with empirical potentials of mean force to evaluate the sequence-structure compatibility. Although the designed sequence has little ordered (secondary) structure in water, circular dichroism and nuclear magnetic resonance data show clear evidence for significant helical content in water/ethylene glycol and in water/methanol mixtures at low temperatures, as well as melting behavior indicative of cooperative folding. We believe that this represents a significant step toward meeting the Paracelsus Challenge.

15/7/80 (Item 5 from file: 73)
DIALOG(R)File 73:EMBASE

06433253 EMBASE No: 1996082453
Application of genetic algorithms to combinatorial synthesis: A computational approach to lead identification and lead optimization
Singh J.; Ator M.A.; Jaeger E.P.; Allen M.P.; Whipple D.A.; Soloweij J.E.; Chowdhary S.; Treasurywala A.M.
Department of Medicinal Chemistry, Sanofi Winthrop, 1250 South Collegeville Road,Collegeville, PA 19426-0900 United States
Journal of the American Chemical Society (J. AM. CHEM. SOC.) (United States) 1996, 118/7 (1669-1676)
CODEN: JACSA ISSN: 0002-7863
DOCUMENT TYPE: Journal; Article
LANGUAGE: ENGLISH SUMMARY LANGUAGE: ENGLISH

A genetic algorithms (GA) based strategy is described for the identification or optimization of active leads. This approach does not require the synthesis and evaluation of huge libraries. Instead it involves iterative generations of smaller sample sets, which are assayed, and the 'experimentally' determined biological response is used as an input for GA to rapidly find better leads. The GA described here has been applied to the identification of potent and selective stromelysin substrates from a combinatorial-based population of 20sup 6 or 64000000 possible hexapeptides. Using GA, we have synthesized less then 300 unique immobilized peptides in a total of five generations to achieve this end. The results show that each successive generation provided better and unique substrates. An additional strategy of utilizing the knowledge gained in each generation in a spin-off SAR activity is described here. Sequences from the first generations were evaluated for stromelysin and collagenase activity to identify stromelysin-selective substrates. GlyProSerThr-TyrThr with Tyr as the Pinf 1' residue is such an example. A number of peptides replacing Tyr with unusual monomers were synthesized and evaluated as stromelysin substrates. This led to the identification of Ser(OBn) as the best and most selective Pinf 1' residue for stromelysin.

15/7/81 (Item 6 from file: 73)
DIALOG(R)File 73:EMBASE

06331890 EMBASE No: 1995365881
Genome sequence comparison and scenarios for gene rearrangements: A test case
Hannenhalli S.; Chappey C.; Koonin E.V.; Pevzner P.A.
Department of Computer Science, University of Southern California,Los Angeles, CA 90089-1113 United States
Genomics (GENOMICS) (United States) 1995, 30/2 (299-311)
CODEN: GNMCE ISSN: 0888-7543
DOCUMENT TYPE: Journal; Article
LANGUAGE: ENGLISH SUMMARY LANGUAGE: ENGLISH

As large portions of related genomes are being sequenced, methods for comparing complete or nearly complete genomes, as opposed to comparing

individual genes, are becoming progressively more important. A major, widespread phenomenon in genome evolution is the rearrangement of genes and gene blocks. There is, however, no consistent method for genome sequence comparison combined with the reconstruction of the evolutionary history of highly rearranged genomes. We developed a schema for genome sequence comparison that includes three successive steps: (i) comparison of all proteins encoded in different genomes and generation of genomic similarity plots; (ii) construction of an alphabet of conserved genes and gene blocks; and (iii) generation of most parsimonious genome rearrangement scenarios. The approach is illustrated by a comparison of the herpesvirus genomes that constitute the largest set of relatively long, complete genome sequences available to date. Herpesviruses have from 70 to about 200 genes; comparison of the amino acid sequences encoded in these genes results in an alphabet of about 30 conserved genes comprising 7 conserved blocks that are rearranged in the genomes of different herpesviruses. Algorithms to analyze rearrangements of multiple genomes were developed and applied to the derivation of most parsimonious scenarios of herpesvirus evolution under different evolutionary models. The developed approaches to genome comparison will be applicable to the comparative analysis of bacterial and eukaryotic genomes as soon as their sequences become available.

15/7/82 (Item 7 from file: 73)
DIALOG(R)File 73:EMBASE

05780456 EMBASE No: 1994196609
Optimal sequence selection in proteins of known structure by simulated evolution
Hellinga H.W.; Richards F.M.
Department of Biochemistry, Duke University Medical Center,Durham, NC 27110 United States
Proceedings of the National Academy of Sciences of the United States of America (PROC. NATL. ACAD. SCI. U. S. A.) (United States) 1994, 91/13 (5803-5807)
CODEN: PNASA ISSN: 0027-8424
DOCUMENT TYPE: Journal; Article
LANGUAGE: ENGLISH SUMMARY LANGUAGE: ENGLISH

Rational design of protein structure requires the identification of optimal sequences to carry out a particular function within a given backbone structure. A general solution to this problem requires that a potential function describing the energy of the system as a function of its atomic coordinates be minimized simultaneously over all available sequences and their three-dimensional atomic configurations. Here we present a method that explicitly minimizes a semiempirical potential function simultaneously in these two spaces, using a simulated annealing approach. The method takes the fixed three-dimensional coordinates of a protein backbone and stochastically generates possible sequences through the introduction of random mutations. The corresponding three-dimensional coordinates are constructed for each sequence by 'redecorating' the backbone coordinates of the original structure with the corresponding side chains. These are then allowed to vary in their structure by random rotations around free torsional angles to generate a stochastic walk in configurational space. We have named this method protein simulated evolution, because, in loose analogy with natural selection, it randomly selects for allowed solutions in the sequence of a protein subject to the 'selective pressure' of a

potential function. Energies predicted by this method for sequences of a small group of residues in the hydrophobic core of the phage lambda cI repressor correlate well with experimentally determined biological activities. This 'genetic selection by computer' approach has potential applications in protein engineering, rational protein design , and structure-based drug discovery.

15/7/83 (Item 8 from file: 73)
DIALOG(R)File 73:EMBASE

02386798 EMBASE No: 1983155809
Mapping the order of DNA restriction fragments
Fitch W.M.; Smith T.F.; Ralph W.W.
Dep. Physiol. Chem., Univ. Wisconsin-Madison, Madison, WI 53706 United States
Gene (GENE) (Netherlands) 1983, 22/1 (19-29)
CODEN: GENED
DOCUMENT TYPE: Journal
LANGUAGE: ENGLISH

A straightforward method was designed for mapping the order of DNA restriction fragments obtained by a double and two single digestions, without the necessity of using a computer or a radioactive label. All possible solutions compatible with a pre-set level of error in the determination of sequence lengths are obtained. The primary assumptions are given, and the appropriate modifications of the algorithm are presented as a function of any assumptions one is unable (or unwilling) to make. Use of the method in connection with end-labeled fragments is also described.

15/7/84 (Item 1 from file: 76)
DIALOG(R)File 76:Life Sciences Collection

02231326 4279295
Structural analysis of a necrogenic strain of cucumber mosaic cucumovirus satellite RNA in planta
Rodriguez Alvarado, G.; Roossinck, M.J.
S. R. Noble Found., Inc., P.O. Box 2180, Ardmore, OK 73402, USA
VIROLOGY vol. 236, no. 1, pp. 155-166 (1997)
ISSN: 0042-6822
DOCUMENT TYPE: Journal article LANGUAGE: ENGLISH
SUBFILE: Virology & AIDS Abstracts; Biochemistry Abstracts 2: Nucleic Acids

Structural studies of plant viral RNA molecules have been based on in vitro chemical and enzymatic modification. That approach, along with mutational analysis, has proven valuable in predicting structural models for some plant viruses such as tobacco mosaic tobamovirus and brome mosaic bromovirus. However, in planta conditions may be dramatically different from those found in vitro. In this study we analyzed the structure of cucumber mosaic cucumovirus satellite RNA (sat RNA) strain D4 in vivo and compared it to the structures found in vitro and in purified virions. Following a methodology developed to determine the structure of 18S rRNA within intact plant tissues, different patterns of adenosine and cytosine modification were found for D4-sat RNA molecules in vivo, in vitro, and

in virions. This chemical probing procedure identifies adenosine and cytosine residues located in unpaired regions of the RNA molecules. Methylation data, a genetic algorithm in the STAR RNA folding program, and sequence alignment comparisons of 78 satellite CMV RNA sequences were used to identify several helical regions located at the 5' and 3' ends of the RNA molecule. Data from previous mutational and sequence comparison studies between satellite RNA strains inducing necrosis in tomato plants and those strains not inducing necrosis allowed us to identify one helix and two tetraloop regions correlating with the necrogenicity syndrome.

15/7/85 (Item 2 from file: 76)
DIALOG(R)File 76:Life Sciences Collection

01982777 3836104
RAPDSIM: Computer programs for modeling the RAPD assay on DNA sequences
Stothard, J.R.
Dep. Zool., Nat. Hist. Mus., Cromwell Rd., London SW7 5BD, UK
J. HERED. vol. 86, pp. 408-409 (1995)
ISSN: 0022-1503
DOCUMENT TYPE: Journal article LANGUAGE: ENGLISH
SUBFILE: Genetics Abstracts

The randomly amplified polymorphic DNA (RAPD) assay is currently used as a tool for genetic mapping and strain determination. Since its inception in 1990, the assay has received a widespread application ill both prokaryotic and eukaryotic molecular genetics and species identification. The novel feature of the RAPD assay is in its application of a single, short oligonucleotide primer to initiate the PCR. This single primer, of arbitrary nucleotide sequence, often has numerous complementary binding sites scattered throughout the template genome. If some of these binding sites are oriented in an inverted or palindromic repeat along the DNA template and are sufficiently close to allow amplification, then ensuing thermal cycling will generate discrete amplification products. Numerous amplification products of varying nucleotide length may be generated within a single RAPD reaction. Agarose or poly-acrylamide gel electrophoresis is subsequently employed to size-fractionate the products, giving rise to a profile analogous to a DNA fingerprint. The fact that RAPDs survey numerous loci in the DNA template makes the method particularly attractive for analysis of genetic distance and phylogeny reconstruction. Although there has been considerable experimental progress in this direction, computational evaluation of the RAPD assay has so far been restricted to simulations involving randomly generated and mutated DNA sequences. In order to complement this study, it would be worthwhile to assess the RAPD assay upon "real" genomes, namely fully sequenced prokaryotes or eukaryotic genes, and evaluate the derived phenetic groupings. In order to facilitate this approach, a collection of menu-driven programs, written in TURBO C (version 1.01, Borland International Inc.). has been designed to simulate the RAPD assay on fully sequenced genes or genomes and randomly generated DNA sequences. The programs can also estimate the F statistic between compared DNA sequences and should prove useful for teachers and researchers wishing to demonstrate to students key features of the RAPD assay without incurring costs from laboratory consumables, design RAPD primers for previously sequenced taxa thereby avoiding prolonged experimental screening of primers, and scan DNA sequences for inverted repeats. The programs run on any IBM-compatible personal computer with the MS-DOS operating system

(version 3.3 or newer). Graphical output from the programs can be carried out in the presence of a Hercules, CGA, EGA, or VGA card. A hard disk for the storage of larger sequences is recommended, though it is not a routine requirement. The executable files are briefly described. RAPDSIM.exe is a menu driven program that is able to simulate the binding of a specific primer against a specified template DNA sequence and reports to the user the number, position, and orientation of annealed primers as well as the nucleotide base composition of the template DNA sequence. RAPDPLOT.exe and RAPDCOM.exe graphically display simulated RAPD fingerprints. RAPDPLOT.exe reports the number and size of the amplified products for a single primer/template combination. RAPDCOM.exe calculates a similarity matrix between four primer/template combinations. FILECON.exe is a file conversion utility designed to convert DNA sequences and from EMBL or GenBank format into a format readable by RAPDSIM.exe. The distribution files contain the programs in execution and source code format, user's instructions, and a DNA sequence data file containing the complete genomic sequence of West Nile Fever Virus in EMBL format. Further information can be found on the distribution diskette by running the batch file INSTALL and viewing the READ.ME file. Individuals interested in the programs should send requests with either a 3.5 double prime or a 5.25 double prime single IBM formatted disk in a self-addressed disk-mailer to the author. The programs will be distributed free of charge. Coupons for return postage required.

013319136
WPI Acc No: 2000-491075/200043
In silico recombinant nucleic acid preparation by genetic algorithm guided gene synthesis involves providing a number of parental character strings, providing oligonucleotides and elongating them
Patent Assignee: MAXYGEN INC (MAXY-N)
Inventor: DEL CARDAYRE S; GUSTAFSSON C ; MINSHULL J; PATTEN P A; SELIFONOV S A ; STEMMER W P C ; TOBIN M
Number of Countries: 091 Number of Patents: 003
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 200042560	A2	20000720	WO 2000US1202	A	20000118	200043	B
AU 200032101	A	20000801	AU 200032101	A	20000118	200054	
EP 1062614	A1	20001227	EP 2000909922	A	20000118	200102	
			WO 2000US1202	A	20000118		

Priority Applications (No Type Date): US 2000416837 A 20000118; US 99116447 A 19990119; US 99118813 A 19990205; US 99118854 A 19990205; US 99141049 A 19990624; US 99408392 A 19990928; US 99408393 A 19990928; US 99416375 A 19991012; US 99416837 A 19991012; US 2000416375 A 20000118
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200042560	A2	E	127	G06F-019/00	

Designated States (National): AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR

IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW
AU 200032101 A G06F-019/00 Based on patent WO 200042560
EP 1062614 A1 E G06F-019/00 Based on patent WO 200042560
Designated States (Regional): AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

Abstract (Basic): WO 200042560 A2

NOVELTY - Preparing recombinant nucleic acid (I) from oligonucleotides which correspond to a set of character string subsequences (SCSS) comprising at least two parental character strings (PCS) corresponding to a number of nucleic acids, is new.

DETAILED DESCRIPTION - Preparing recombinant nucleic acid (I) by aligning for maximum identity a number of parental character strings (PCS) corresponding to a number of nucleic acids, defining a set of character string subsequences (SCSS) comprising at least two of the PCS, providing a set of oligonucleotides corresponding to the SCSS and then annealing and elongating one or more oligonucleotides with polymerase or ligating at least two with ligase to produce (I).

INDEPENDENT CLAIMS are also included for the following:

(1) preparing character strings (CS) by providing PCS encoding a polynucleotide or polypeptide , providing a set of oligonucleotide character strings of preselected length that encode a number of single-stranded oligonucleotide sequence comprising sequence fragments of PCS and its complement and creating a set of derivatives of parental sequence comprising sequence variant strings, a set of multiple mutations with one mutation per variant string;

(2) a library prepared by the above said method;

(3) facilitating recombination between two or more divergent nucleic acids by aligning PCS corresponding to divergent nucleic acids, identifying regions of sequence identity and regions of sequence diversity, defining a diplomat CS which is intermediate in PCS, synthesizing at least a portion of the diplomat sequence to produce a diplomat nucleic acid and recombining a mixture of parental nucleic acid and diplomat nucleic acid;

(4) a mixture of selected nucleic acids produced by the above said method;

(5) generating and recombining nucleic acids by inputting a number of amino acid sequence character strings (ASCS) into a digital system, reverse translating ASCS in the digital system to a number of nucleic acid character strings which are species codon biased in a selected expression host and with optimized sequence similarity between a number of nucleic acid character strings and synthesizing one or more oligonucleotides from one or more reverse translated nucleic acid sequences;

(6) optimizing activity of a nucleic acid by parameterizing a number of nucleic acids or proteins to provide a set of multidimensional datapoints, extrapolating one or more postulated multidimensional datapoint from the set of multidimensional datapoints and converting the postulated multidimensional datapoint to a new CS corresponding to a postulated nucleic acid or protein ;

(7) providing a library of recombinant nucleic acids which is enriched for a sequence of interest and selecting the library by producing an initial library of at least about 106 recombinant nucleic acids, comprising at least about 105 different non-identical units, hybridizing the library to one or more population of nucleic acids that correspond to one or more subsequences in the different library units;

(8) the enriched library produced by the above said method;

(9) generating a library of biological polymers by generating a diverse population of CS in a computer, which in turn are generated by alteration of pre-existing CS, synthesizing the diverse population of CS in which diverse population comprises the library of biological polymers; and

(10) an integrated system comprising a computer having a first data set comprising a first CS, a second data set comprising a second CS, software for aligning the first and second CSs, software for performing a genetic operation on the first or second CS, an output file comprising a third data set comprising a third CS, the third CS comprising CS subsequences from the first and second CSs, and an oligonucleotide sequence output file comprising a plurality of overlapping oligonucleotide sequences corresponding to third CS.

USE - The method is useful for rapid evolution of nucleic acids in vitro and in vivo and provides for generation of encoded molecules with new and/or improved properties. Proteins and nucleic acids of industrial, agricultural and therapeutic importance can be created or improved through DNA shuffling procedures.

ADVANTAGE - Physical access to genes or organisms is not required as sequence information is used for design and selection of oligo. Extensive sequence information is provided and sequences from inaccessible, non cultivable organisms can also be used. Sequences from pathogens without actual handling of pathogens and all type sequences including damaged and incomplete genes are amenable to this method. All genetic operators and crossovers can be fully and independently controlled in a reproducible fashion removing human error and variability from physical experiments with DNA manipulations. Sequences with frame-shift mutations are eliminated or fixed. Wild type parents do not contaminate derivative libraries with multiple redundant parental molecules.

pp; 127 DwgNo 0/15

Derwent Class: B04; D16; T01

International Patent Class (Main): G06F-019/00

15/7/87 (Item 2 from file: 351)

DIALOG(R)File 351:Derwent WPI

011713706 **Image available**

WPI Acc No: 1998-130616/199812

Method for screening peptide(s) with high physiological activity - by determining activity, dropping least active, changing sequence in remaining peptide(s) and cycling through this sequence to obtain peptide(s) with high activity

Patent Assignee: KARUBE M (KARU-I); YAMANOUCHI PHARM CO LTD (YAMA)

Inventor: KARUBE M; YOKOBAYASHI Y

Number of Countries: 021 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 9804580	A1	19980205	WO 97JP2535	A	19970723	199812	B
EP 916677	A1	19990519	EP 97932982	A	19970723	199924	
			WO 97JP2535	A	19970723		
JP 10508678	X	19991019	WO 97JP2535	A	19970723	200001	
			JP 98508678	A	19970723		

Priority Applications (No Type Date): JP 96198096 A 19960726
Patent Details:
Patent No Kind Lan Pg Main IPC Filing Notes
WO 9804580 A1 J 19 C07K-001/00
Designated States (National): CA JP US
Designated States (Regional): AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE
JP 10508678 X C07K-001/00 Based on patent WO 9804580
EP 916677 A1 E C07K-001/00 Based on patent WO 9804580
Designated States (Regional): AT BE CH DE DK ES FI FR GB GR IE IT LI LU NL PT SE

Abstract (Basic): WO 9804580 A
Method for screening a group of peptides (preferably 5-20 amino acid residues in length) in order to obtain peptides with an enhanced physiological activity comprises:
(a) synthesising a number of different peptides;
(b) determining the particular physiological activity to be enhanced, for each peptide;
(c) selecting those with the highest activity, and
(d) processing the sequences of the selected peptides using a genetic algorithm computer program which:
(i) exchanges one or more amino acid residues between pairs of peptides (crossover) and/or
(ii) substitutes amino acid residues in a peptide for different ones (mutation) (preferably with a mutation frequency of about 3% of the total number of residues).
The altered peptide sequences obtained by this means are then synthesised and the cycle (1)-(4) repeated.
It is found that the average activity of the peptide set rises with each generation and eventually peptides with a highly enhanced activity are obtained.
USE - Peptides with a high activity (e.g. as enzyme inhibitors) are obtained by a relatively fast and effective method to give products suitable for use in the drug, foodstuff and other industries.
Dwg.4/4
Derwent Class: B04; D16
International Patent Class (Main): C07K-001/00
International Patent Class (Additional): G06F-015/18

15/7/88 (Item 1 from file: 202)
DIALOG(R)File 202:Information Science Abs.

00191831 9501831
ISA Document Number in Printed Publication: 9501770
Optimization and performance analysis of a massively parallel dynamic programming algorithm for RNA secondary structure prediction.
Document Type: Journal Article
Author (Affiliation): Shapiro, B.A. (National Institutes of Health, Frederick, MD); Chen, J.-H.; Busse, T.; Navetta, J.; Kasprzak, W.; Maizel, J.
Country of Affiliation: United States
Journal: International Journal of Supercomputer Applications and High Performance Computing
Publication Language(s): English

Source: Vol. 9 Issue 1 p. 29-39 Spr 1995 13

An optimized and parallelized form of a dynamic programming algorithm capable of generating optimal and suboptimal RNA secondary structures is presented. Implementation of this algorithm on a MasPar MP-2 with 16K processors is shown to perform extremely well for very large nucleic acid sequences such as HIV (AIDS) and Rhinovirus (common cold). These sequences are, respectively, 9,128 and 7,208 nucleotides in length. By taking advantage of the parallel nature of MasPar and also optimizing the communication requirements the implementation essentially reduced the algorithm from order O(n sup 3 to O(n sup 2). This reduction is complexity enables us to fold large RNA sequences in reasonable amounts of time. This capability has proven to be a valuable tool in studying the molecular structure and biological function of molecules this large.

?logoff hold